

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 14:27:26 ; Search time 35.8 Seconds
(without alignments)
2935.079 Million cell updates/sec

Title: US-09-840-746-1

Perfect score: 4859

Sequence: 1 MSQETVSRSVAPMRGGEIT.....FPGQYNPSFISDESRDRDY 946

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802:*
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4798	98.7	1481	17	AAW03740
2	4798	98.7	1481	20	AAW03740
3	4798	98.7	1481	20	AAW03740
4	4798	98.7	1481	21	AAW03740
5	4798	98.7	1481	21	AAW03740
6	4798	98.7	1481	21	AAW03740
7	1280	26.3	235	22	AAW03740
8	689	14.2	156	22	AAW03740
9	689	14.2	156	22	AAW03740
10	689	14.2	156	22	AAW03740
11	689	14.2	156	22	AAW03740

12	523	10.8	121	22	AAU16941	Human novel secret
13	377	7.8	234	22	AAU37120	Staphylococcus aur
14	326	6.7	957	21	AAU59288	Human MUCII polype
15	326	6.7	957	22	AAW24513	C900P predicted am
16	325	6.7	849	17	AAW06725	FLOI protein, invo
17	324	6.7	1257	15	AAW46627	Neurocan core prot
18	317	6.5	1795	22	ABR69806	Drosophila melanog
19	310	6.4	560	22	AAU37464	Staphylococcus aur
20	307.5	6.3	2586	22	ABR66878	Drosophila melanog
21	307	6.3	688	22	ABB30137	Peptide #2788 enco
22	307	6.3	688	22	ABB35307	Peptide #2813 enco
23	307	6.3	688	22	ABB20749	Protein #2748 enco
24	307	6.3	688	22	AAW56138	Human brain expres
25	307	6.3	688	22	AAW68511	Human bone marrow
26	307	6.3	688	22	AAW16315	Peptide #2749 enco
27	307	6.3	688	22	AAW28810	Peptide #2847 enco
28	307	6.3	688	22	AAW04053	Peptide #2735 enco
29	302	6.2	1237	21	AAW81609	Streptococcus pneu
30	301.5	6.2	1532	21	AAW40945	Human ORFX ORF709
31	299	6.2	2870	21	AAW95559	Caenorhabditis ele
32	299	6.2	3178	21	AAW95556	Caenorhabditis ele
33	290.5	6.0	894	15	AAW47578	Flocculation prote
34	290.5	6.0	894	15	AAW58754	S. cerevisiae FLOI
35	284	5.8	502	22	AAU34408	Staphylococcus aur
36	280	5.8	862	15	AAW60563	Yeast 2.6 kb agglu
37	279.5	5.8	1537	15	AAW60562	Yeast 4.7 kb agglu
38	277	5.7	5179	22	AAW24516	C899P predicted am
39	276.5	5.7	2409	12	AAW12609	Versican. Homo sa
40	269.5	5.5	2858	22	ABB71150	Drosophila melanog
41	269.5	5.5	3060	22	ABB58064	Drosophila melanog
42	268.5	5.5	571	22	AAW27242	Human EXMAD-20 SEQ
43	264.5	5.4	881	22	AAW70752	S. cerevisiae apopt
44	263.5	5.4	750	20	AAW05477	C. albicans Rbt1 p
45	263	5.4	2471	20	AAW06816	Human Notch2 (hum

ALIGNMENTS

RESULT 1	
AAW03740	
ID	AAW03740 standard; Protein; 1481 AA.
XX	
AC	AAW03740;
XX	
DT	20-NOV-1996 (first entry)
XX	
DE	rchd528 gene product.
XX	
KW	Cardiovascular disease; differential expression; target gene;
KW	pathway gene; fingerprint gene; atherosclerosis; ischaemia;
KW	reperfusion; hypertension; restenosis; arterial inflammation;
KW	vector; antibody; diagnosis; gene therapy; drug screening;
KW	rchd528 gene.
XX	
OS	Homo sapiens.
XX	
PH	Key
FT	Peptide
FT	/label= Sig_peptide
FT	1089..1122
FT	/label= Extracellular domain
FT	/note= "comprises an epidermal growth factor repeat motif"
FT	
FT	Modified-site
FT	1140..1151
FT	/label= Asn_hydroxylation_site
FT	/note= "asparagine hydroxylation site consensus sequence"
FT	
FT	Domain
FT	1348..1370
FT	/label= Transmembrane_domain
XX	
PN	WO9624604-A1.

PD 15-AUG-1996.
XX PF 09-FEB-1996; 96WO-US01883.
XX PR 07-JUN-1995; 95US-0485573.
PR 10-FEB-1995; 95US-0386844.
XX PA (MILL-) MILLENNIUM PHARM INC.
XX PI Falb DA;
XX DR WPI; 1996-384391/38.
DR N-PSDB; AAT36035.
XX PT New genes differentially expressed in cardiovascular disease - and
PT related vectors, host cells, proteins and antibodies, for diagnosis,
PT monitoring, treatment and drug screening
XX PS Example 9; Fig 31A-D; 200pp; English.
XX CC The amino acid sequence (AAW03740) was deduced of the protein encoded
CC by the novel human rchd528 gene (AAT36035). rchd528 is very highly
CC expressed in the heart. The rchd528 gene is up-regulated in
CC endothelial cells subjected to shear stress or exposed to oestrogen.
CC It is one of 8 novel genes (see also AAT36029-36) found to be
CC differentially expressed in cardiovascular disease (CVD), and is
CC useful in methods for the diagnosis and treatment of CVD. Assays
CC using cells expressing the novel genes can be used to identify cpds.
CC exhibiting anti-CVD activity.
XX SQ Sequence 1481 AA;

Query Match 98.7%; Score 4798; DB 17; Length 1481;
Best Local Similarity 90.3%; Pred. No. 1.5e-257;
Matches 945; Conservative 1; Mismatches 0; Indels 100; Gaps 1;

QY 1 MSQTETVRSVAPMRGEITAHLLNTSTTSADVTCSSASYPEGVNASVLTFQSDSTVQS 60
Db 436 msqtetvrsvapmrggeitahllntsttsadvtcssasypegvnasvltfqsdstvqs 495
QY 61 GCSHTALGDRSYSESSTSSSESLNSAPRGERS----- 94
Db 496 ggshtaalgdrsyssesstsseslnsaprgerstledsrepdgaigdsanaedrsgv 555
QY 95 ----- 94
Db 556 pslgthlatvtngertllrsvltnmtsttsmettsgeagspaampqetegaslhvrvtdm 615
QY 95 -----IAGISVGQVRGTAIEQRTSSDHTDHTYLLSTFTKGERALLSITDNS 140
Db 616 glvrselaassaigvagisygqvrtaieqrtsdhtdhtylstftkgerallsitdns 675
QY 141 SSSDIVESSSYIKINSSHSEYSSFSHAQTERSNTSSYDGEYAQPTESPVLHSTNLP 200
Db 676 sssdivesssyikinsnhseysfsshaqtersnssydgeyaqptespvlhtsnlps 735
QY 201 YPTIIMPNTSVVLDTAERFVSSSSSSSSSSSSSGPPPLPLPSVSQSHLFSSTILPST 260
Db 736 yptiimpntsvvldtaerfvsssssssssssgppplpsvsqshlfsstlpsst 795
QY 261 RASVHLLKSTSDATPWSSSPSPPLVSLTWTSTGAPLSVSGSTTLPQSSSTPVLPRARETPV 320
Db 796 rasvhllkstdatpwssspplvslttstgapsvsgsttlpqssstpvlpraretpv 855
QY 321 TSFQSTMTSFMPLHSSOTADILKSQSTPHQKVITESKSPSLVSLPTSTKAVTTNSPL 380
Db 856 tsfqtstmtsfmtplhssotadilksqstphqkviteskspslvslptstkvttncspl 915
QY 381 PPSLTSSTSTQTLPATSTNIAQMSPTFTTTILKTSQPLMTTPGTLSTASLVTGPIAVQT 440
Db 916 ppsltsststqtlpatstnlqmsptftttlilktspmltptgtlstaalvtgpiavqt 975

QY 441 TAGKQLSLTHPEILVPOISTEGGISTERNRIVDATTGLIPLTSVPTSAKEMTTKLGVTA 500
Db 976 tagkqlslthpeilvpqisteggisternrvivdattgllpltsvptsakemttklgvt 1035
QY 501 EYSPASRLGTSPTSPQTVVSTAEADLPKSAFVQSGSTQSPPTLSSASVNSCAVNPCL 560
Db 1036 eyspasrlgtspsptvtvstaeadlpkstatfavqsgstqpttllssasvnscavpnc 1095
QY 561 HNGECVADNTSRGYHCRCPSPSWQDDCSVDVNECLSNPCPSTATCNTTQGSFICKCPVG 620
Db 1096 hngecvadntsrghyhcrcpsspwwgddcsvdvneclsnpcpstatcnttqgsfickcpvg 1155
QY 621 QLEKGICNLVRTFVTEFKLRTFLNTTVEKHSDLQEVENEITKTLNMCFSALPSYIRSTV 680
Db 1156 qlekgicnlvrtfvtefkrlrtflnttvekhsdlqeveneitklnmcfsalpsyirstv 1215
QY 681 HASRESNAVVISIQTTFSLASNVTLFDLARMQKVCVNSCKSSAEVCCOLLGSRQIFRAGS 740
Db 1216 hasresnavvisiqttfslasnvtlfdlarmqkvcvnsckssaevcqllgqrifrag 1275
QY 741 LCKRKSPEDKDTSICTDLGVALCCKSGVFOFNKMDHSCRACEDGYRLENETCMSCPF 800
Db 1276 lckrkspedkdtictdlgvalccksgvfofnkmdhscracedyrilenetcmscpf 1335
QY 801 GLGLNCGNPYQLITVIVIAAAGGGLLILGIALIVTCCRKNKNDISKLIKFSGDFQMSPY 860
Db 1336 glglnccgncpyqlitvviaaagglllilgialivtccrknkndisklifiksgdfqmspy 1395
QY 861 AEYPKNPRSQEWGREATIEHENGSTKNLLQMTDVIYVYSPTSVRNPELERNGLYPAYTGLPG 920
Db 1396 aeypknprsqewgreatiehengstknllqmtdvyyyspvsrnpelernglypaytg 1455
QY 921 SRHSCIFPGYGNPFSIDESRRRDYF 946
Db 1456 srhscifpgdynpsfidesrrrdyf 1481

RESULT 2
AAW95160
ID AAW95160 standard; Protein; 1481 AA.
XX AC AAW95160;
XX DT 24-MAY-1999 (first entry)
XX DE Amino acid sequence of rchd528 gene product.
XX KW Fingerprinting gene; rchd502; transmembrane protein; cardiovascular;
KW fingerprint/target gene; up-regulated; endothelial cell; shear-stress;
XX KW atherosclerosis; ischemia; reperfusion; hypertension; restenosis.
XX OS Homo sapiens.
XX PN US5882925-A.
XX PD 16-MAR-1999.
XX PF 09-FEB-1996; 96US-0599654.
XX PR 09-FEB-1996; 96US-0599654.
PR 10-FEB-1995; 95US-0386844.
PR 07-JUN-1995; 95US-0485573.
XX PA (MILL-) MILLENNIUM PHARM INC.
XX PI Falb DA;
XX DR WPI; 1999-214071/18.
DR N-PSDB; AAX26246.
XX PT New polynucleotides consisting of residues 1-1929 of the rchd502
PT gene - are differentially expressed in cardiovascular disease
PT states, and can therefore be used to treat and diagnose

PT cardiovascular diseases
 XX Disclosure; Fig 30A; 121pp; English.
 PS
 CC The invention relates to a rchd502 target/fingerprint gene encoding a
 CC transmembrane protein. The invention provides cDNAs contained in plasmids
 CC pFCHD502SF (ATCC 69981) and pFCHD502SJ (ATCC 69982) that encode the
 CC rchd502 polypeptide, and are differentially expressed in cardiovascular
 CC disease states. Cultured genetically engineered host cell containing the
 CC rchd502 polynucleotides in operative association with a nucleotide
 CC regulatory element are used for producing a polypeptide rchd502 gene
 CC product. Identifying that the fingerprint/target gene rchd502 is
 CC differentially expressed (up-regulated) by endothelial cells subjected
 CC to shear-stress, provides a tool for the diagnosis and treatment of
 CC cardiovascular disease e.g. atherosclerosis, ischemia/reperfusion,
 CC hypertension, restenosis. The fingerprint gene is useful for testing the
 CC efficacy of candidate drugs in basic research and in clinical trials and
 CC or imaging of a diseased cardiovascular tissue. The gene may also be
 CC used in screening for ligands of target gene product receptor domains, as
 CC well as antagonists of the ligand-receptor interaction.
 XX
 SQ Sequence 1481 AA;

Query Match 98.7%; Score 4798; DB 20; Length 1481;
 Best Local Similarity 90.3%; Pred. No. 1.5e-257;
 Matches 945; Conservative 1; Mismatches 0; Indels 100; Gaps 1;

QY 1 MSQFTVSRVAPMRGGEITAHWLLTNTSTTSADVTGSSASYPEGVNASVLTQFSDSTVQS 60
 DB 436 msqtetvsvrvmrggeitahwllntsttsadvtgssasypegvnasvltqfstdstvsq 495
 QY 61 GGSHTALGDRYSSESSTSSSLSNSAPRGERS----- 94
 DB 496 ggshtaigdrysessstsseslnsaprgerstledrepqgalgdsanaedrtsgv 555
 QY 95 ----- 94
 DB 556 pslgthlatvtgngertlrsvtlntsmsttsgeagspaaampqetegaslhvnrddm 615
 QY 95 -----IAGISYQGVRGTAIEQRTSSDHTDHTYLSSTFTKGERALLSITDNS 140
 DB 616 glvrsllaasalgvagisvgvrgtaieqrtssdhtdhtylsstftkgerallsidns 675
 QY 141 SSSDIVESSRYIKINSNSHSEYSSFHAQTERSNISYDGEYAOPTESPVLHTSNLPS 200
 DB 676 sssdivessryiklnssshseyssfhaqtersnissydgeyaoptespvlhtsnlps 735
 QY 201 YTPPINPNTSVLDTDAEFVSDSSSSSSSSSSSSSGPPLPLPSVQSOSHHLPSSILPST 260
 DB 736 ytpinmpntsvltdaeefvsssssssssssssgpplpplpsvqsqshlrfssilpst 795
 QY 261 RASVHLKSTSDASTPWSSSPPLPVSLTSTTSAPLSVSQTTLPOSSSTPVLPRARETPV 320
 DB 796 rasvhlkstdastpwwsspsplpvsltsttsaplsvsqtlpqssstpvpraretpv 855
 QY 321 TSFOTSTWTFMTMLHSQADLKSQSTPQROEKVITESKSPSLVSLPSTESTKAVTTNSPL 380
 DB 856 tsfotstwmfmlhsqadlksqstpqroekviteskpslvsllpstedskavttnspl 915
 QY 381 PPSLTESSTEQTLPATSTNLQAQMSPTTTTLTKTSQPLMTPTGTLSTASIVTGPIAVQT 440
 DB 916 ppsltessteqtlpatstnlaqmspttttltktsqplmtptgtlssasivtgpiavqt 975
 QY 441 TAGKQLSHTPEILVPOISTEGGISTERNRNVIVDATTGLIPLTSVPTSAKEMTTKLGVTA 500
 DB 976 tagqlslhtpelvpoisteggisternrvivdatgllptsvtsakemttklgvt 1035
 QY 501 EYSPASRLGTSPTQPTVSTVSTADLAPKATFAVQSTQSPPTLLSSASVNSCAVNPCL 560
 DB 1036 eyspasrlgtsptqptvstvstadlapkatsfavqstqspptllssasvnscavpcl 1095
 QY 561 HNGECVADNTSRGYHCRCPSPWQGDGCSVDVNECLSNPCPSTATCNNTQGSFICKCPGY 620

DB 1096 hngecvadntsrghcrppswqgdgcsvdvneclsnpcpstatcnntqgsfickcpvy 1155
 QY 621 QLEKIGCNLVRTFTEFKLRTFTNTVEKHSIDLOEVENEITKTLMCMCFSAALPSYIRSTV 680
 DB 1156 qlekgicnlvrtfvtelkrlftnttvekhsidloeveneitktlmcmcfalsalpsyrstv 1215
 QY 681 HASRESNAVVISLQTTSLASNVTLFDLADRMOKCVNSCKSSAEVCCOLLGSSORTRFRAGS 740
 DB 1216 hasresnavvislqttslasnvtfldladrmaqkcvnsckssaevcqlgssqrtrfrags 1275
 QY 741 LCKRKSPECCKDTSICTDGLGVALCQCKSGYFQFNKMDHSCRACEDGYRLNENETCMSCPF 800
 DB 1276 lckrkspecckdtsictldgvalcqcksgyfqfnkmdhscracedgyrlenetcmscpf 1335
 QY 801 GLGGLGCGNPYQLTVVIAAGGGLLLILGIALIVTCCRNKNNDISKLIKFSGDFQMSPY 860
 DB 1336 glgglncgnyqlitvviaaaggllililgialivtccrknkndisklifsqdfqmspy 1395
 QY 861 AEYKPNRSPQEWGREATEMHENGSTKNLQMTDYYVYSPTSVRNPELERNGLYPAYTGLPG 920
 DB 1396 aeypknprsqewgreatemhengstknllqmtdyyvysvtnpelernglypaytgpg 1455
 QY 921 SRHSCIFPGQYNPSFISDESRRDYF 946
 DB 1456 srhscifpgqynpsfidesrrdyf 1481

RESULT 3
 AAW89299
 ID AAW89299 standard; Protein; 1481 AA.
 XX
 AC AAW89299;
 XX
 DT 11-MAR-1999 (first entry)
 XX
 DE Human rchd528 protein.
 XX
 KW Human; cardiovascular disease; atherosclerosis; ischaemia; restenosis;
 KW reperfusion; hypertension; arterial inflammation; diagnosis; rchd528.
 XX
 OS Homo sapiens.
 XX
 PN US5849578-A.
 XX
 PD 15-DEC-1998.
 XX
 PF 15-MAR-1996; 96US-0616844.
 XX
 PR 09-FEB-1996; 96US-0599654.
 PR 10-FEB-1995; 95US-0386844.
 PR 07-JUN-1995; 95US-0458873.
 PR 15-MAR-1996; 96US-0616844.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Falb DA;
 XX
 XX WPI; 1999-069743/06.
 DR N-PSDB; AAV81825.
 DR
 XX
 PT DNA encoding rchd528 polypeptide - associated with cardiovascular
 PT disease
 XX
 XX Claim 1; Fig 30; 122pp; English.
 PS
 CC The present sequence represents rchd528 protein. A method has been
 CC developed for producing the rchd528 gene product. The present invention
 CC describes methods and compositions for the treatment and diagnosis of
 CC cardiovascular diseases, including: atherosclerosis; ischaemia;
 CC restenosis; reperfusion; hypertension; and arterial inflammation.
 XX
 SQ Sequence 1481 AA;

```

Query Match          98.7%;   Score 4798;   DB 20;   Length 1481;
Best Local Similarity 90.3%;   Pred. No. 1.5e-257;
Matches 945;   Conservative 1;   Mismatches 0;   Indels 100;   Gaps 1;

QY 1 MSQTETVSRVAPMRGGEITAHLLTNTSTTSADVTGSSASYPEGVNASVLTFQFSDSTVQS 60
Db 436 msqtetvsvapmrgeitahlltntsttsadvtgssasypegvnasvltfqfstdstvsq 495
QY 61 GGSHTALGDRSYFSSSTSSSLSNAPRGERS----- 94
Db 496 ggshtaigdrsyessstsseslnsaprgerstledsrepqgalgdssanaadrtsgv 555
QY 95 ----- 94
Db 556 pslgthlatvgtngertlrvltlntsmsttagaagspaaapqetegaslhvntddm 615
QY 95 -----IAGISYQVGRGTAEIQRSTSDHTPDHTYLSSTFTFKGERALLSIFDINS 140
Db 616 glvrsllaasalgvagisygvrgtaieqrtsdhdtdtylsatftckgerallsicdms 675
QY 141 SSSDIVESSTYIKISNSHSEYSSFSHAQTERSNISSYDGEVAQPTESPVLHTSNLPS 200
Db 676 sssdivessscyikinsnsheysfsfnagqcrsnlsydgayaqptespvlhtsnlps 735
QY 201 YTPTINMPNTSVVLDTDAEFVSDSSSSSSSSSSSGPPLPLPSVSQSHHLFSSILPST 260
Db 736 ytptinmpntsvvldtdaeivsdsssssssssssgpplpipsvsgshhlfsilpst 795
QY 261 RASVHLKSTSDASTPWSSSPPLPVSLTWTSTAPLSVSQTTPLPQSSSTPVLPRARTPV 320
Db 796 rasvhlktsdstpwwsspsplpvslttstaplsvsqttlpqssstpvlprrartpv 855
QY 321 TSFQTSMTWTFMHLHSSQADLKSOQTPHOEKVITESKPSLSVLSTPTESTKAVTTSNPL 380
Db 856 tsfqtstmtwtfmhlhssqadlksqtphoekviteskpslsvlptestkavttnspl 915
QY 381 PPSLTESSTEQTLPATSTNLQAQSPFTTTLKTSQPLMTTPGTLSTSLVTPGPIAVQT 440
Db 916 ppsltessteqtlpatstnlqagsptfttlktsqplmttptgtlslstslvtpgpiavqt 975
QY 441 TACKQLSLTHPELILVPOISTEGGISTERNVIVDATTGLIPLISVPPTSAREMTTKLGVT 500
Db 976 tagkqlslthpelilvpqisteggisternvividattgllpltsvptsakemtktklgvt 1035
QY 501 EYSPASRLSGTSPQRTVYSTADLAPKSAFVQSQSTQPTTLSSASVNSCAVNPCL 560
Db 1036 eyspasrlsgtspqrtvystaedlapksatfavqsgstqpttlssasvnscavpncpl 1095
QY 561 HNGECVADNTSRGHCRCPPSWGGDCSDVDNECLSNPCPSTATCNTNQSGFICKCPVGY 620
Db 1096 hngecvadntsrghrcppswggdcsvdneclsnpcpstatcnnntggsfickcpvgy 1155
QY 621 OLEKGINLVRTFTVEFKLRTFLNTTVEKHSLOQEVENEITKTLNMCFSALPSYIRSTV 680
Db 1156 qlekginlvrtftvefklrtflnttvekhsdlqeveneitktlnmcfsalpsyirstv 1215
QY 681 HASRESNAVVISLQTTTSLASNTVLFOLADRMOKCVNSCKSSAEVCCOLGQRRIFRAGS 740
Db 1216 hasresnavvislqtttflslasntvlfoladrmqcvnsckssaevoqlgqrrifrag 1275
QY 741 LCKRKSPCEDKDTISICPDLGVALCOCKSGYFQFNKMDHSCRACEDGYRLNETCMSCPF 800
Db 1276 lckrkspcedkdtisicpdlgvalcqcksgyfqfnkmdhscracedyrlnetcmscpf 1335
QY 801 GLGLNCGNPYLITVVVIAAGGELLILGIALIVTCCRKNKNDISKLIFFKSGDFQMSPY 860
Db 1336 glglncgnpylitvvviaaggellilgialivtccrknkndisklifksgdfqmspy 1395
QY 861 AEPKPNRQSEGWEGREATEMHENGSTKNLLQMTDVIYSPTSVRNPFLERNGLYPAYTCLPG 920
Db 1396 aepkpnrsqsewqegreatemhengstknllqmtdviysptsvrnpflernglypaytclpg 1455

```

```

QY 921 SRHSICFPQVNPSPFISDESRRRDYF 946
Db 1456 srhscifpgqynpsfidsesrrrdyf 1481

RESULT 4
AAB19626
ID AAB19626 standard; Protein; 1481 AA.
XX AAB19626;
XX 05-FEB-2001 (first entry)
XX Human rchd528 gene protein product.
XX Human; rchd528; differential expression; HUVEC; shear stress;
KW endothelial cell; cardiovascular disease; inflammation;
KW atherosclerosis; antiinflammatory; antiatherosclerotic; diagnosis;
KW therapy.
XX Homo sapiens.
XX Key Location/Qualifiers
FT Peptide 5..28 /label= Signal_peptide
FT Protein 29..1431 /label= Mature_protein
FT Domain 1348..1370 /label= Transmembrane_domain
FT Region 1089..1122 /note= "epidermal growth factor repeat motif"
FT Modified-site 1140..1151 /note= "Asn is hydroxylated"
XX US6124433-A.
XX 26-SEP-2000.
XX 06-OCT-1997; 97US-0944496.
XX 09-FEB-1996; 96US-0599654.
XX 10-FEB-1995; 95US-0386844.
XX 07-JUN-1995; 95US-0485573.
XX (BGHM ) BRIGHAM & WOMENS HOSPITAL.
XX (MILL-) MILLENNIUM PHARM INC.
XX Gimbrone MA, Falb DA;
XX WPI; 2000-611017/58.
XX N-PSDB; AAA88578.
XX Novel isolated rchd502 polypeptides, differentially expressed in
XX response to endothelial cell shear stress, used for diagnosis,
XX monitoring clinical trails, and treating cardiovascular diseases such
XX as ischemia -
XX Example 9.2; Fig 30A-I; 123pp; English.
XX
CC The present sequence is that of the predicted polypeptide encoded
CC by the newly identified human rchd528 gene (see AAA88578). This gene
CC is differentially expressed (up-regulated) in response to endothelial
CC cell shear stress, and is not induced by interleukin-1. The
CC predicted protein is a novel G protein coupled receptor including 7
CC transmembrane domains. It shows 40% identity to the angiotensin II
CC receptor. Up-regulation of the rchd528 gene in a disease state may
CC reflect a protective role for the gene product in responding to
CC disease. Alternatively, it may have a causative or exacerbating
CC effect on the disease state. Modulation of rchd528 gene expression,
CC or the activity of its gene product, may provide a protective
CC effect. Knowledge of the gene and its protein product will provide
CC for drugs with greater specificity for the treatment of inflammation

```


CC and atherosclerosis. rchd528 is 1 of 8 novel human genes of the
CC invention (see A988576-83) characterised as being differentially
CC expressed in cardiovascular disease states, and which are of
CC diagnostic or therapeutic use.

XX	Sequence	1481 AA;	
SQ	Query Match	98.7%; Score 4798; DB 21; Length 1481;	
	Best Local Similarity	90.3%; Pred. No. 1.5e-257;	
	Matches	945; Conservative 1; Mismatches 0; Indels 100; Gaps 1;	
QY	1	MSQETVSRVAPMRGGEITAHMLLTNSTTSADVTSASYPEGVNASVLTFQSDSTVQS	60
DB	436	msqtetvsrvapmrgeitahwllntsttsadvtsasypegvnasvltqfsdstvqs	495
QY	61	GGSHALGDRYSSESSSTSESLSNSAPRGERS-----	94
DB	496	ggshtalgrtysesssstsseslnssaprgertstledsrepqgalgssanaedrtsgv	555
QY	95	-----	94
DB	556	pslghthlatvtngentlrsvltntsmsttsgeagspaaampqetegaslhvntddm	615
QY	95	-----IAGISYGOVRGTAEIORTSSDHTDHTYLSSTFTKGERALLSITDNS	140
DB	616	glvrslaaasalgvagisvqrgtaldqrdtsdhcdhtyilstftkgerallsitdms	675
QY	141	SSSDIVESSTYIKISNSHSEYSEFSFHAQTERSNISYDGEYAQPSTPVLHTSNLPS	200
DB	676	sssdivesstsyikisnshseysefsfhaqternsisyydgyaqpstevpvlhtsnlps	735
QY	201	YTPTINPNTSVLDTDAEFVSDSSSSSSSSSSSSSGPPLPLPVSQSHHLSIILPST	260
DB	736	ytptinmpntsvvldtadaefvsssssssssssssgpplpipvsqshhlfssilpst	795
QY	261	RASVHLKSTSDASTPWSSPPLPVSLTSTAPLSVSOTTLPOSSSTPVLPRARETPV	320
DB	796	rasvhlkstdastpwspsplpvsltstaplsvsqtlpqssstpvpraretpv	855
QY	321	TSFQTSMTSFMTHLSSQATADLKQSTPHQEVITESKPSLSVLPSTESTKAVTNSPL	380
DB	856	tsfqtstmtsfmthlssqatadlkqstphqeviteskpslvspltestkavtntspl	915
QY	381	PPSLTSTSTEQTLPATSTNLQAQSPPTTTILKTSQPLMTPTGTLSTASLVTCPIAVQT	440
DB	916	ppsltststeqtlpatstnlqaqspptfttlktsqplmtptgtltsaslvtpgiavqt	975
QY	441	TAGKQLSLTHPEILVPOISTEGGISTERNRIVDATTGLIPLTSVPTSAKEMTTKLGVTA	500
DB	976	tagkqlslthpeilvpisteggisternrvivdattgllpltsvplsakemttklgvta	1035
QY	501	EYSPASRLGTSPTQTVVSTAEADLAPKATFAVQSTQSTPTLSSASVNSCAVNPCL	560
DB	1036	eyspasrlgtspsptqtvvstaeadlapkatsfavqstqstptllssasvnscavpnc1	1095
QY	561	HNGECVADNTRSRYHRCPCPSWQDDCSVDVNECLSNPCSTATCNTQCSFFICKCPVG	620
DB	1096	hngecvadntrsryhrcpcpswqddcsvdvneclsnpcstacntqgsffickcpvg	1155
QY	621	QLKKGICNLVRTFTVEFKLRTFLNTTVEKHSDIQEYENETTKLNNCFSAIPSYIRSTV	680
DB	1156	qlkgicnlvrtftvefklrtflnttvekhdsdiqevenettklmmcfalsalpsyirstv	1215
QY	681	HASRESNAVVISIQTTFSLASNTVFLDLADRMQKVCVNSCKSSAEVFCOLLGSRQRIFRAGS	740
DB	1216	hasresnavvisiqttsfslasntvflldlrmqkvcvnsckssaevcqlgssrqrifrag	1275
QY	741	LCRKRSPECCKDYSICTDLGVALCQCKSGYFQFNKMDHSCRACEDGYRIENETCMSCPF	800
DB	1276	lckrkspecckdysictdlgvalcqcksgyfgfnkmdhscracedgyrienetcmcpf	1335
QY	801	GLGLNGCNGPYQLITTVIAAAGGGLLLILGIALIVTCCRNKNNDISKLIKSGDFQMSPY	860

DB	1336	glgngcngpyqlitvviaaagglililgialivtccrknknndiskliksgdfqmspy	1395
QY	861	AEYPKPNRSQEWGREATIMHENGSTKNLQMTDYYYSPTSVRNPPELRNGLYPAYTGLPG	920
DB	1396	aeypkpnrsqewgrealeimhengstknllqmtdvyyptsrvnpelernglypaytglpg	1455
QY	921	SRHSCIFPGQYNPSFISDESRRRDYF	946
DB	1456	srhscifpgqynpsfisderrrdyf	1481
RESULT	5		
AA	AA78508		
ID	AA78508	standard; Protein; 1481 AA.	
XX	AA78508;		
XX	AC		
XX	05-MAY-2000	(first entry)	
XX			
DE		Amino acid sequence of cardiovascular disease associated protein rchd528.	
XX			
KW		Differentially expressed; cardiovascular disease; atherosclerosis;	
KW		ischaemia; reperfusion; hypertension; restenosis; arterial inflammation;	
KW		rchd528; transmembrane protein.	
XX			
OS		Homo sapiens.	
XX			
PN		US6020463-A.	
XX			
PD		01-FEB-2000.	
XX			
PF		06-OCT-1997; 9705-0944423.	
XX			
PR		09-FEB-1996; 9605-0599654.	
PR		10-FEB-1995; 9505-0386844.	
PR		07-JUN-1995; 9505-0485573.	
XX			
PA		(BCHM) BRIGHAM & WOMENS HOSPITAL.	
PA		(MILL-) MILLENNIUM PHARM INC.	
XX			
PI		Gimbrone MA, Falb DA;	
DR		WPI: 2000-146911/13.	
DR		N-PSDB; AA289796.	
XX			
PT		Marker proteins for the diagnosis of cardiovascular diseases such as	
PT		atherosclerosis and hypertension, comprising peptide sequences derived	
PT		from the rchd523 transmembrane protein -	
XX			
PS		Examples; Fig 20; 121pp; English.	
XX			
CC		This sequence represents the human rchd528 amino acid sequence. This	
CC		sequence is related to the rchd523 transmembrane polypeptide which is	
CC		encoded by cDNA contained in the plasmid pfchd523. The rchd523 protein is	
CC		differentially expressed in diseased cells compared to healthy cells. The	
CC		rchd523 protein may be used as a marker protein for the diagnosis of	
CC		cardiovascular diseases including atherosclerosis, ischaemia,	
CC		reperfusion, hypertension, restenosis and arterial inflammation. rchd523	
CC		peptides may be used as antigens in the production of antibodies specific	
CC		for rchd523. The anti-rchd523 antibodies may then be used in diagnostic	
CC		assays to quantitate rchd523 peptides in samples.	
XX			
SQ		Sequence 1481 AA;	

Query Match	98.7%;	Score 4798;	DB 21;	Length 1481;
Best Local Similarity	90.3%;	Pred. No. 1.5e-257;		
Matches	945;	Conservative 1;	Mismatches 0;	Indels 100; Gaps 1;
QY	1	MSQETVSRVAPMRGGEITAHMLLTNSTTSADVTSASYPEGVNASVLTFQSDSTVQS	60	
Db	436	msqtetvsrvapmrgeitahwllntsttsadvtsasypegvnasvltqfsdstvqs	495	

Qy	61	GGSHALGDASYSESSTSSSELN\$APRGERS-----94
Db	496	ggshalgdrtsyessstssseln\$aprgerstledrpggqalgdssanaedrsgv555
Qy	95	-----94
Db	556	pslgtlatvtngertlrsvtlntsmsttsgeagapaaampqetegaslhvnvtddm615
Qy	95	-----IAGISGYQRGTAI\$BQRTSSDHTHYLSSTFTYGERALL\$ITDNS140 :
Db	616	glvrslaagaalvagiygvgrgtaleqrctssdhtdhtylsstftkgerallsitdns675
Qy	141	SSSDIVESSFYIKINSRHSYSVSPHQAOTRSNISVGDEYAQPSTESPLVTSLNLP\$200
Db	676	ssedivesatsyklennshesyafshaqtarsnisgydeyaqpstespyltlnlp\$735
Qy	201	YTPTINPNTSVVDNDAEFVSODSSSSSSSSSGSGGPPPLPSVSOSHHLFS\$ILPE\$T260
Db	736	ytpcinmpntsvvltdaeufvdssassssssssgpplplpsavsgghlf\$asilpst795
Qy	261	RASVHLLKTS\$DASTPWSSPGLPVSLTLT\$T\$APL\$YSO\$TTLPQSSSTPYLPRARETPV320
Db	796	rasvhllktsdstpwssppplpvsltltsaplsvsqttlpqssstpylpraretpv855
Qy	321	T\$QT\$MTSFMTHL\$NSTADLK\$OSTPBHQEKVITESK\$PSLV\$LPTTESTKAVT\$TN\$PL380
Db	856	tsfqtsmtsfmhlhsqtdaliksgstphqeikvitesk\$pslv\$lpttestkavt\$tnspl915
Qy	381	P\$SLTES\$E\$TLPAT\$TNLAQ\$MPTFTTILKT\$QPLW\$TTPGTL\$S\$TASL\$VTGP\$IAVO\$T440
Db	916	ppls1tessteqtlpaestnlagnspfcftcliktsgqlmtcpgtlsscaslvtgpiavqt975
Qy	441	TAGKQL\$U\$THEILVPQISTEGGI\$TERNRV\$VDATTGLIPLT\$VPTS\$AKEMT\$TKLG\$VTA500
Db	976	tagkqialthpeilvpqisteggisternrvivdatgtgliptsyptsakemtclkvta1035
Qy	501	EY\$PASRL\$TG\$TSPOTT\$V\$TAE\$DLAPK\$ATFAVQS\$TQ\$PTL\$SSAS\$VNC\$AVNPCL560
Db	1036	eypasrsrltgspqdtvvsdaedlapksatfvqsatqpttlsssasvncavnpc1095
Qy	561	HNGECVADNTSRGYHCRCPP\$WQGDDCSVDVNECLSNPC\$STATC\$NNQ\$GFICKCPVGY620
Db	1096	hngecvadntsryhcrcpwwgddcsvdvneclsnpcpstatcnntqgsfickcpvy1155
Qy	621	OLEKGICNLVRTVTEFKLRFLNTVY\$KH\$DLQEVENETKTLNMCF\$ALPYIR\$STV680
Db	1156	qlekigicnlvrtvtefkrlrflntvtyekhdsldqeveneitktlnmcfsalpsyirstv1215
Qy	681	HASRESNAVVISLQ\$TTF\$LIASNVTLF\$DLADRMQKVNSCK\$SAEVCOLLGSORIRFRAG\$740
Db	1216	hasresnavvislqtifslasnvtlfdladrmqkvnscksaevcollgsorirfrags1275
Qy	741	LCKRK\$PEC\$DKDT\$ICTDLDGVALCOCK\$SGYEQFNKM\$PH\$CRACBDGYRLENETCM\$CPF800
Db	1276	lckrkspcdkdtictdlgdvalcocksgyfqnkmphscracedgyrlenetcmcpf1335
Qy	801	GLGLLGCNPNQYLITVVIAAAGG\$LLLLIGIALIVTCRKNKNDISKILIFK\$SGDFQMS\$PY860
Db	1336	glglgcncnpnylitvviaaaggllligialivtcrcrnknndiskilfksgdfqmspy1395
Qy	861	AEPYKPNR\$QEWGREAIEMHENG\$TNKLQMTDVY\$PT\$VRNPELRNGLYPAYTGLPG920
Db	1396	aepkpnrsqewgreaiemhengstknlqmtdvysptvrnpelelnglypaytblpg1455
Qy	921	SRHSCIFPGQYNPFISDSERRDYF946
Db	1456	srhscifpgqynpfisdesrrdyf1481

RESULT 6

AAV68447

ID AAY68447 standard; Protein; 1481 AA.

[illegible]

XX	02-AUG-2001.	PR	29-SEP-2000;	2000US-236367P.
PD		PR	29-SEP-2000;	2000US-236368P.
XX		PR	29-SEP-2000;	2000US-236369P.
XX	17-JAN-2001;	PR	29-SEP-2000;	2000US-236370P.
PF	2001WO-US01349.	PR	02-OCT-2000;	2000US-236802P.
XX		PR	02-OCT-2000;	2000US-237037P.
PR	31-JAN-2000;	PR	02-OCT-2000;	2000US-237038P.
PR	04-FEB-2000;	PR	02-OCT-2000;	2000US-237039P.
PR	24-FEB-2000;	PR	02-OCT-2000;	2000US-237040P.
PR	03-MAR-2000;	PR	02-OCT-2000;	2000US-239335P.
PR	16-MAR-2000;	PR	02-OCT-2000;	2000US-239337P.
PR	17-MAR-2000;	PR	13-OCT-2000;	2000US-239937P.
PR		PR	13-OCT-2000;	2000US-240960P.
PR	18-APR-2000;	PR	20-OCT-2000;	2000US-241221P.
PR	19-MAY-2000;	PR	20-OCT-2000;	2000US-241785P.
PR	07-JUN-2000;	PR	20-OCT-2000;	2000US-241786P.
PR	28-JUN-2000;	PR	20-OCT-2000;	2000US-241787P.
PR	30-JUN-2000;	PR	20-OCT-2000;	2000US-241808P.
PR	07-JUL-2000;	PR	20-OCT-2000;	2000US-241809P.
PR	07-JUL-2000;	PR	20-OCT-2000;	2000US-241826P.
PR	11-JUL-2000;	PR	01-NOV-2000;	2000US-244617P.
PR	11-JUL-2000;	PR	08-NOV-2000;	2000US-246474P.
PR	14-JUL-2000;	PR	08-NOV-2000;	2000US-246475P.
PR	26-JUL-2000;	PR	08-NOV-2000;	2000US-246476P.
PR		PR	08-NOV-2000;	2000US-246477P.
PR	14-AUG-2000;	PR	08-NOV-2000;	2000US-246478P.
PR	14-AUG-2000;	PR	08-NOV-2000;	2000US-246523P.
PR	14-AUG-2000;	PR	08-NOV-2000;	2000US-246523P.
PR	14-AUG-2000;	PR	08-NOV-2000;	2000US-246525P.
PR	14-AUG-2000;	PR	08-NOV-2000;	2000US-246526P.
PR	14-AUG-2000;	PR	08-NOV-2000;	2000US-246527P.
PR	14-AUG-2000;	PR	08-NOV-2000;	2000US-246528P.
PR	14-AUG-2000;	PR	08-NOV-2000;	2000US-246532P.
PR	14-AUG-2000;	PR	08-NOV-2000;	2000US-246609P.
PR	14-AUG-2000;	PR	08-NOV-2000;	2000US-246610P.
PR	14-AUG-2000;	PR	08-NOV-2000;	2000US-246611P.
PR	18-AUG-2000;	PR	08-NOV-2000;	2000US-246613P.
PR	22-AUG-2000;	PR	17-NOV-2000;	2000US-249207P.
PR	22-AUG-2000;	PR	17-NOV-2000;	2000US-249208P.
PR	22-AUG-2000;	PR	17-NOV-2000;	2000US-249209P.
PR	30-AUG-2000;	PR	17-NOV-2000;	2000US-249210P.
PR		PR	17-NOV-2000;	2000US-249211P.
PR	01-SEP-2000;	PR	17-NOV-2000;	2000US-249212P.
PR	01-SEP-2000;	PR	17-NOV-2000;	2000US-249213P.
PR	01-SEP-2000;	PR	17-NOV-2000;	2000US-249214P.
PR	01-SEP-2000;	PR	17-NOV-2000;	2000US-249215P.
PR	05-SEP-2000;	PR	17-NOV-2000;	2000US-249216P.
PR	05-SEP-2000;	PR	17-NOV-2000;	2000US-249217P.
PR	06-SEP-2000;	PR	17-NOV-2000;	2000US-249218P.
PR	06-SEP-2000;	PR	17-NOV-2000;	2000US-249244P.
PR	08-SEP-2000;	PR	17-NOV-2000;	2000US-249245P.
PR	08-SEP-2000;	PR	17-NOV-2000;	2000US-249264P.
PR	08-SEP-2000;	PR	17-NOV-2000;	2000US-249265P.
PR	08-SEP-2000;	PR	17-NOV-2000;	2000US-249297P.
PR	08-SEP-2000;	PR	17-NOV-2000;	2000US-249299P.
PR	08-SEP-2000;	PR	17-NOV-2000;	2000US-249300P.
PR	08-SEP-2000;	PR	01-DEC-2000;	2000US-250160P.
PR	12-SEP-2000;	PR	01-DEC-2000;	2000US-250391P.
PR	14-SEP-2000;	PR	05-DEC-2000;	2000US-251030P.
PR	14-SEP-2000;	PR	05-DEC-2000;	2000US-251988P.
PR	14-SEP-2000;	PR	05-DEC-2000;	2000US-256719P.
PR	14-SEP-2000;	PR	06-DEC-2000;	2000US-251479P.
PR	14-SEP-2000;	PR	08-DEC-2000;	2000US-251856P.
PR	14-SEP-2000;	PR	08-DEC-2000;	2000US-251868P.
PR	14-SEP-2000;	PR	08-DEC-2000;	2000US-251869P.
PR	21-SEP-2000;	PR	08-DEC-2000;	2000US-251989P.
PR	21-SEP-2000;	PR	08-DEC-2000;	2000US-251990P.
PR	25-SEP-2000;	PR	11-DEC-2000;	2000US-254097P.
PR	25-SEP-2000;	PR	05-JAN-2001;	2001US-2559678P.
PR	26-SEP-2000;	XX		
PR	27-SEP-2000;	XX	(HUMA-) HUMAN GENOME SCI INC.	
PR	27-SEP-2000;	XX	Rosen CA, Barash SC, Ruben SM;	
PR	29-SEP-2000;	XX		
PR		XX		

DR WPI: 2001-476161/51.
DR N-PSDB; ABA06457.
XX
PT Isolated nucleic acid molecule encoding an inflammation-associated
PT polypeptide is used in preventing, treating or ameliorating a medical
condition -
XX
XX
XX Claim 11; SEQ ID NO: 543; 859pp + Sequence Listing; English.
XX
XX The present invention provides human cDNAs, proteins and related genomic
CC DNAs. These can be used in the treatment of neural, immune system,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders and inflammation. The present sequence
CC is a protein of the invention.
XX
XX
SQ Sequence 156 AA;

Query Match 14.2%; Score 689; DB 22; Length 156;
Best Local Similarity 85.0%; Pred. No. 5.2e-31;
Matches 130; Conservative 3; Mismatches 16; Indels 4; Gaps 1;

QY 496 LGVTAEYSPASRLGTSPTQTTWSTAEDLAPKSAFVQSSQTQSPPTLSSASVNSCA 555
I : |||||
Db 4 laiqstspasrlgtsptqttwstaedlapksatfavqstqspptlssasvnsca 63
|||||

QY 556 VNPCLHNGECVADNTRGVCRCPPSPQWQGDGCSVDVNECLSNPCPSATCNNTQGSFICK 615
|||||
Db 64 vnpclhngecvadtsgyhcrcppsqqgddcsvdvneclsnpcpsatcnnctqgxxick 123
|||||

QY 616 CPVGYQLEKIGICNLVRFVTFEFLKRTFLNTTV 648
|||||
Db 124 cpvgylkexicnlgk----rlxlftrltti 152
|||||

RESULT 9
AAU18104
ID AAU18104 standard; Protein; 156 AA.
XX
AC AAU18104;
XX
XX
DT 21-NOV-2001 (first entry)
XX
XX Novel human uterine motility-association polypeptide #11.
XX
XX Human; uterine motility-association disorder; uterus; pregnancy;
KW labour; menstrual cycle; gene therapy.
XX
XX Homo sapiens.
XX
XX WO200155201-A1.
XX
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-0501317.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180828.
PR 24-FEB-2000; 2000US-0184564.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.

PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0237935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI: 2001-476222/51.
XX N-PSDB; AAS26918.
XX
PT Novel polypeptides and polynucleotides useful as diagnostic reagents to
PT diagnose diseases or disorders associated with aberrant expression or
PT activity of polypeptides, for treating blood clotting disorder,
PT haemophilia
XX
PS Claim 11; SEQ ID No 254; 601pp; English.
XX
XX The invention relates to isolated nucleic acid molecules and their
XX encoded secreted proteins. The nucleic acids and proteins are used to
XX prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX rabbits, goats, horses, cats, dogs, chickens or sheep. They
XX are also used in diagnosing a pathological condition or susceptibility
XX to a pathological condition. Antibodies to the proteins can also
XX be used in alleviating symptoms associated with the disorders and in
XX diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
XX immunosorbant assays (ELISA). Disorders which are diagnosed or treated
XX include autoimmune diseases e.g. rheumatoid arthritis,
XX hyperproliferative disorders e.g. neoplasms of the breast or liver,
XX cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
XX e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
XX Alzheimer's disease, infections caused by bacteria, viruses and fungi
XX and ocular disorders e.g. corneal infection, and many other
XX disorders listed in the specification. The polypeptides can also
XX be used to aid wound healing and epithelial cell proliferation, to
XX prevent skin aging due to sunburn, to maintain organs before
XX transplantation, for supporting cell culture of primary tissues, to
XX regenerate tissues and in chemotaxis. The polypeptides can also be used

PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239355.
 PR 13-OCT-2000; 2000US-0239337.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246533.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 XX WPI; 2001-476222/51.
 DR N-PSDB; AAS26846.
 XX
 XX Novel polypeptides and polynucleotides useful as diagnostic reagents to
 PT diagnose diseases or disorders associated with aberrant expression or
 PT activity of polypeptides, for treating blood clotting disorder,
 PT haemophilia
 XX
 PS Claim 11; SEQ ID No 182; 601pp; English.
 XX
 CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. Antibodies to the proteins can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
 CC and ocular disorders e.g. corneal infection, and many other
 CC disorders listed in the specification. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present

Query Match 10.8%; Score 523; DB 22; Length 121;
 Best Local Similarity 89.8%; Pred. No. 5.9e-22;

```
Matches 97; Conservative 1; Mismatches 10; Indels 0; Gaps 0;
QY 510 GTSPPQTIVVSTAEADLPKATFAVGSSTQSPPTLSSASVNCVAVNCPCLHNGECVADN 569
Db 1 gtsppqgtvvstaeadlpkatsatfavgsstgptlssasvncvavncpclhngecvadvn 60
QY 570 TSGYHCRCPSPWGDGDCSDVNVCLSNPCPSTATCNCNTQGSFICKCP 617
Db 61 tsrgyhcrppswgddcdvncvncclsnpcpstatwqygsillylqmp 108
RESULT 13
AAU37120
ID AAU37120 standard; Protein; 2344 AA.
XX AAU37120;
XX
DT 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #1290.
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
WP: 2001-611495/70.
DR N-PSDB; AAS54979.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 12713; 51lpp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, themselves and the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
```

```
SQ Sequence 2344 AA;
Query Match 7.8%; Score 377; DB 22; Length 2344;
Best Local Similarity 24.0%; Pred. No. 3.8e-12;
Matches 195; Conservative 149; Mismatches 350; Indels 120; Gaps 24;
QY 2 SQTETVRSVAPMRGGEITAHWLLTNSTTSADV-----TCSSASYPEGVNASVLTQFS 54
Db 1054 sdsdskslslstsgsgts-----tststssvrmsesqsgsmstsgdsstsfsts 1108
QY 55 DSTVQSGSHALGDRSYSESSSTSSSESLNS---SAPRGRSTAGISYGVGRCTAEQ 110
Db 1109 dtsdsksaasta-sseisqsvstsgvstsgstslstlensertstmsdstslstseed 1167
QY 111 RTSSDHTDHTYLSSTFTKGERALLSITDNSSSDI-VESSTSYIKIS---NSSHEYSSF 166
Db 1168 stsdstsdtsiseaisgsestsislssesnstsdskasafisessestseasl 1227
QY 167 SHAOTERSNISSYDGEYAQPSTE-----SPVLHTSNLPS-YPTINMPNTSVVLDTD- 217
Db 1228 sgstsdstsdtsnsesgtsstslsnstsgasistsgsaststvkseavstslstst 1287
QY 218 AEFVSDSSSSSSSSSSSGPPLPL-PSVQSHHLFSSILPSTRASVHLLKSTSDA--- 273
Db 1288 stslsdstslstsdtsdgsknsleamstsdsksrkesisastslsgstsesg 1347
QY 274 STPWSSSPS---PLPVSLTTSAPLSVSQT-TLPOSSSTPVLPRAR-----EPVTSF 323
Db 1348 stsseskdstsmalsmqstsgstsvstseslstdstslsasnmqgvdnsasq 1407
QY 324 QTSTWTSFMTMLHSSQ-TADLKQSQSTPHQKVITESKPSLSVLPTESTKAVTINSPLPP 382
Db 1408 saststsdtsesdsgstsytsqseststslstsdtsksksgststas 1467
QY 383 SLTESSTEQTLPATSTNLQMSPTFTTILKTSOPLMTTPGTLSTASLVTGPIAVQTTA 442
Db 1468 sgseesdsgkstsesksestslsdts-----tsnsgsaststllnsaaseed 1523
QY 443 GKQLSLTHPEILVPOISTEGGISTERNVIVDATTGLIPLTSVPT-SAKEMTTKLGVTAE 501
Db 1524 ssstslsdtsaasmqsgsedsgstslsngqststslrmstiasesvsestsesgstse 1583
QY 502 YSPASRSIGTSPS-PQTTVVSTAEADLPKATFAVGSSTQSPPTLSSASV----- 551
Db 1584 stsesdtsdtsldsgstsrsts---aaggsaststsdtsrstaaststltdsgs 1640
QY 552 -----NSCAVNPCLHNGECVADNTSRGYHCRCPSPWGDGDCSDVNVCLSNPCP-STAT 604
Db 1641 mslststsvsdtsldsvsdts-----dststsgsmaasislstdt 1687
QY 605 CNNTQGSFICKPVGYLEKICMLVRTFTVTFEFLKRTFLNTVTEKHSDLOEVEWEITKT 664
Db 1688 ststasevmsasi-----sdsqsmesvnds 1714
QY 665 LNMCFALPSYIRTVTHASRESNAVVISLQTF-----SLASNVTILFDLADRMQCVNSCK 720
Db 1715 evsesnesdskmsgstsvsdsgslsvtslkrkesvsesislsgsgmsdstsvstsd 1774
QY 721 SSAEVCQLLQSRRIFRAGSLCKRKSPECDKDT 754
Db 1775 sslsvstsqrsesvseesdtsldskstsgststs 1808
RESULT 14
AAV59288
ID AAV59288 standard; Protein; 957 AA.
XX
AC AAV59288;
XX
DT 25-APR-2000 (first entry)
XX
DE Human MUC11 polypeptide.
```

XX Mucin; MUC11; MUC12; human; chromosome 7q22; epithelial inflammation;
 KW Crohn's disease; ulcerative colitis; asthma; chronic bronchitis;
 KW colorectal cancer; cystic fibrosis; inflammatory bowel disease;
 KW breast cancer.
 XX Homo sapiens.
 XX Key Location/Qualifiers
 FH Modified-site 68
 FT /note: "potential N-glycosylation site"
 PN W0200004142-A1.
 XX 27-JAN-2000.
 XX 16-JUL-1999; 99WO-AU00579.
 XX 16-JUL-1998; 98AU-0004708.
 XX (COUN-) COUNCIL QUEENSLAND INST MEDICAL RES.
 PA (ORDE-) ORDER OF SISTERS OF MERCY IN QUEENSLAND.
 XX Williams SJ, Antalis TM, McGuckin MA, Gotley DC;
 XX WPI; 2000-182416/16.
 XX N-PSDB; AA258824.
 XX Novel MUC nucleic acid corresponding to mucin gene, useful for treating
 PT associated disease conditions e.g. colorectal, breast cancer, cystic
 PT fibrosis and inflammatory bowel disease -
 XX Claim 10; Page 83-88; 103pp; English.
 XX The invention provides mucin genes (MUC11 and MUC12) located on human
 CC chromosome 7q22. The mucin genes or its portion is used in detecting
 CC polymorphism, mutation, deletion, truncation and expansion in the gene
 CC or its gene transcript. Pharmaceutical compositions and gene therapy
 CC constructs comprising the mucin genes are used for treating disease
 CC conditions associated with aberrant Mucin expression, altered properties
 CC of mucus or epithelial inflammatory processes involving Mucins like
 CC Crohn's disease, ulcerative colitis, asthma, chronic bronchitis and
 CC colorectal cancer, cystic fibrosis, inflammatory bowel disease and breast
 CC cancer. The mucin genes and the polypeptides are used for determining
 CC these diseases or their predisposition. The MUC11 and MUC12 polypeptides
 CC are used for preparing antagonist and antibodies. The present sequence
 CC represents the human MUC11 polypeptide.
 XX Sequence 957 AA;
 SQ
 Query Match 6.7%; Score 326; DB 21; Length 957;
 Best Local Similarity 26.1%; Pred. No. 7.8e-10;
 Matches 168; Conservative 84; Mismatches 243; Indels 148; Gaps 28;
 QY 28 STTSADVTGSSASYPGVNASHVLFQFSDSTVQSG-----GSHPALG--DRSYSESSSTS 79
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 342 sttsgvseestshspgthtfaftdttptglsrhtstshspgtdtllpasttt 401
 QY 80 ---SSSLNSSAPRGERSIAGISGVQVRGAIEORTSSDHTD----HTYLSSTTKGERA 132
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 402 sgpsgeatshspgtdta-lspgststafsggesttfhspgsthttl----- 450
 QY 133 LLSLTDNSSSDIYESSTYKISNHSSEYSSPSHA-----QTERSNISSYDGE-YA 184
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 451 ---fpdsttsglveast---rvhsstgsprttlspasstspglqgestafqtphastht 504
 QY 185 QPST-----ESPVLHSTNLPSYPTINMPNTSVV-----LDTDAEFVSDSSSSSS 230
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 505 tpsstpatapveesttyhrs--psstptthfpasssttshgskstfthspdsagttps 562
 QY 231 SSSSSSSGPPPLPLPSVQSHLHFFSILPSTRASVHLKSR-----SDASTPWSSSP-SPL 284
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 563 sahattag-----rgeattsrspgsteittlpgsttptglseastfyssprsppt 613
 QY 285 ----PVSL-----TTSTSAPLSVSQTTLPSSSSPPVLPRARETPVTQFTSTMT-- 329
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 614 tllspasmtslgvgeeststsrspgsthststspasttllpglseesttvysspgstetv 673
 QY 330 ---SFTMTLHSSQTADLKSQSTPHQEKVITESKPSLSLPTSTEST----KAVTNSPLPP 382
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 674 fprstttsvrgepttthfhrpasthtttlftedstts--glteestafgspastqtgipa 731
 QY 383 SLT-----SSSTEQTL-PATSTNL-----AQMSPFTTTTILKTSQPLM 419
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 732 tlttadlgeesttffsssgstgttllsparsttsglvgestsrspststetttlpgsp-- 789
 QY 420 TTPGTLSTASLVTCPTAVQTTAGKQLSLTHPEILLVPOISTEGGISTERNRIVVDATGL 479
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 790 ttpalseksttfftsprspdat-----lspatttssgvsee-----sstshs 831
 QY 480 IPLTSVPTSAKEMTTKLGVTAEYSPASRSLSGTSPSPQTTVVSTARDLAPKSAFVQSSST 539
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 832 qpsthttaftdttsttsglsqepkthshsqgsteatlspgattasslqgqsttftf-sapg 890
 QY 540 QSPPTLSSSASVNSCAVNP-----CLHNGECVADNTSRG 573
 DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
 891 dtettllpddtitsglveastpshststgslhtltpasstsaag 933
 RESULT 15
 AAM24513
 ID AAM24513 standard; Protein; 957 AA.
 XX AAM24513;
 XX 12-OCT-2001 (first entry)
 XX C900P predicted amino acid sequence.
 DE Human; immunotherapy; diagnosis; colon cancer; colon tumour;
 KW Immunogenic; gene therapy; vaccine; colonic cancer.
 XX Homo sapiens.
 OS W0200149716-A2.
 PN 12-JUL-2001.
 XX 29-DEC-2000; 2000WO-US35596.
 XX 30-DEC-1999; 99US-0476296.
 PR 10-JAN-2000; 2000US-0480321.
 PR 15-FEB-2000; 2000US-0504629.
 PR 06-MAR-2000; 2000US-0519444.
 PR 19-MAY-2000; 2000US-0575251.
 PR 29-JUN-2000; 2000US-0609448.
 PR 28-AUG-2000; 2000US-0649811.
 XX (CORI-) CORIXA CORP.
 XX Xu J, Lodes MJ, Secrist H, Benson DR, Meagher MJ, Stolck JA;
 PI King GE, Wang T, Jiang Y;
 XX WPI; 2001-441847/47.
 XX Colon tumor associated proteins and nucleic acids useful for the
 PT prevention, diagnosis and treatment of colonic cancer -
 XX Claim 2; Page 437-440; 472pp; English.
 XX The present invention describes colon tumour associated proteins (I) and
 CC the polynucleotides (II) that encode them. (I) have cytostatic activity.
 CC (I) and (II) can be used in gene therapy and vaccine production. (I) and
 CC (II) may be used in the prevention, diagnosis and treatment of diseases
 CC associated with inappropriate colon tumour associated protein (TCAP)

expression, such as colonic cancer. For example, (I) and (II) may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of TCAPs by expressing inactive proteins or to supplement the patients own production of them. Additionally, (II) may be used to produce the TCAP proteins, by inserting the nucleic acids into a host cell culturing the cell to express the protein. (II) and its complementary sequences may also be used as DNA probes in diagnostic polymerase chain reaction (PCR) and hybridisation assays to detect and quantitate the presence of similar nucleic acids in samples, and therefore which patients may be in need of restorative therapy. (I) may also be used as antigens in the production of antibodies against TCAPs and in assays to identify modulators of TCAP expression and activity. Anti-(I) antibodies and antagonists may also be used to down regulate TCAP expression and activity. The anti-(I) antibodies may also be used as diagnostic agents for detecting the presence of TCAPs in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). AA128460 to AA129512 and AA244494 to AA244523 represent nucleotide and amino acid sequences given in the exemplification of the present invention.

Sequence	957 AA;
SQ	

Query Match	Score	DB 22:	Length
Best Local Similarity	6.7%	326;	957;
Matches 168; Conservative	26.1%;	Pred. NO. 7.8e-10;	
Matches 148; Gaps	84;	Mismatches 243;	
Indels	148;		

QY 28 STTSADVTGSSASYPEGVNASVLTFPSDSTVQSG-----GSHTALG--DRSYSESSSTS 79
||||| : | : :: | ||| | ||:: | | : :||:
||||| : | : :: | ||| | ||:: | | : :||:

```

Db      342 sttsgvseettshsrpggthtaftpdsttppglrhattshspgsttdtllpasatt 401
Qy      80 ---SSELSNSAPPGRSIAISYGQVRGWAIBQRTSSDHTD---HTVLSSTFTKGERA 132
Db      402 sgpsqettshspgstcdta-lpggsttalsfgqestthfshpggsthttl----- 450

```

```
QY 133 LL$ITDSSSDIVESSTSYKISNSSHSEVSSFSA-----QTERSNISSVDGE-YA 184
      |::|||::|| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::| ::::|
Db 451 ---fpd$ttssgivaest---rvhsstgsprtlspasstsqqlqgestafqtbpastht 504
```

QY	185	QPST-----	ESPVLHTSNLPSTPTINPNTSVV----	LDTDAEFFVSDSSSSSS	230
			:		
Db	505	tpstptatapveesttyhrs--	psstptthfpassttsqhskestifhsspdasqtpps	562	

QY 231 SSSSSSGPPLPLPSVSQSHHLFSSLLPSTRASVHLKST-----SDASTPWSSSP-SPL 284
| : | : || : | | : | | : | | : | | : | | : | | : | | : | | : | | : | | : |
Dd 563 sahsttsg-----rqesttsrpsqsteitlpqastttbpqlseattffvssprsp 613

```

QY 285 ----PVSL-----TTSTAPLSVSQTLPQSSSTPVLPRARETPVTSFQTSTMT-- 329
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 614 tllspasmts|qvgeesttsrsqposthstvspsasttbpalssesttvvsssoqstetv 673

```

QY 330 ---SFMTMLHSSQTADLKSGSTPHQEKVITKSKSPSLVSLPSTEST----KAVTTNSPLPP 382
| | : : : : :
pb 674 forsttttsvrgaeoottfhsroasthttlftedgitts--glteastaafasnaatatale 731

```

QY      383 SLT-----ESTEQTL-PATSNL-----AQMSPFTTTLKTSQPLM 419
      :||      || || || || || || || || || || || || || || || ||
pb      732 tlttadlqaestttfnsssgstgttignarststglnvactnsgtsttttlogen- 780

```

Qy 420 TTPTGLSSTASLVTGPIAVQTAGKQLSLTHPEILVQISTEGGISTERNRVIVDATTGL 479

QY 480 IPLTSVPTSAKEMTTKLGVTAEYSPASRSLGTSPPOTTVVSTAEDLAPKSATFAVQSST 539

QY 540 QSPITLSSASVNCVNP-----CLHNGECVADNTSRG 573

22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 10

Search completed: July 23, 2002, 14:30:06
Job time: 160 sec

This Page Blank (uspio)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 14:27:06 ; Search time 18.31 Seconds
(without alignments)
1261.967 Million cell updates/sec

Title: US-09-840-746-1
Perfect score: 4859
Sequence: 1 MSQETVSRVAPMRGGEIT.....FPGQVNPFSIDSERRRDYF 946

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/6C_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/6D_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4798	98.7	1481	2	US-08-616-844-40
2	4798	98.7	1481	2	US-08-599-654-40
3	4798	98.7	1481	3	US-08-944-868A-40
4	4798	98.7	1481	3	US-08-944-423A-40
5	4798	98.7	1481	3	US-08-944-496-40
6	328	6.8	1257	1	US-08-340-428B-49
7	290.5	6.0	894	3	US-08-362-525-22
8	290.5	6.0	894	3	US-08-971-692-15
9	284.5	5.9	1537	1	US-08-325-267A-2
10	280	5.8	862	1	US-08-325-267A-4
11	279	5.7	2523	1	US-08-185-432-18
12	276.5	5.7	2409	6	5180808-2
13	263.5	5.4	750	4	US-09-165-239A-4
14	254	5.2	2471	1	US-08-185-432-16
15	254	5.2	2471	1	US-08-083-590A-19
16	254	5.2	2471	3	US-08-532-384-19
17	240.5	4.9	2703	1	US-08-185-432-19
18	240	4.9	1193	2	US-08-400-159-10
19	240	4.9	1193	3	US-08-611-729A-10
20	236	4.9	2476	4	US-08-276-967-2
21	233	4.8	1523	4	US-09-182-024A-2
22	231.5	4.8	1010	4	US-08-882-046-7
23	231.5	4.8	1036	4	US-09-068-740A-6
24	231.5	4.8	1187	4	US-09-068-740A-7
25	231.5	4.8	1218	3	US-08-400-159-6
26	231.5	4.8	1218	3	US-08-611-729A-6
27	231.5	4.8	1218	4	US-08-882-046-2

28	231.5	4.8	1218	4	US-09-214-278-7	Sequence 7, Appli
29	231.5	4.8	1218	4	US-09-068-740A-11	Sequence 11, Appli
30	231	4.8	729	3	US-08-872-855-8	Sequence 8, Appli
31	231	4.8	2556	1	US-08-083-590A-20	Sequence 20, Appli
32	231	4.8	2556	3	US-08-532-384-20	Sequence 20, Appli
33	230	4.7	907	3	US-08-783-774-2	Sequence 2, Appli
34	230	4.7	907	5	PCT-US95-04611A-19	Sequence 19, Appli
35	226	4.7	1219	4	US-08-882-046-5	Sequence 5, Appli
36	226	4.7	2556	1	US-08-185-432-17	Sequence 17, Appli
37	225.5	4.6	728	4	US-08-981-392-2	Sequence 2, Appli
38	222.5	4.6	1248	4	US-08-882-046-6	Sequence 6, Appli
39	218.5	4.5	1055	4	US-09-214-278-2	Sequence 2, Appli
40	218.5	4.5	1212	4	US-09-214-278-3	Sequence 3, Appli
41	218.5	4.5	1238	4	US-09-214-278-5	Sequence 5, Appli
42	218.5	4.5	1257	3	US-08-611-729A-8	Sequence 8, Appli
43	217	4.5	1721	3	US-08-928-361B-6	Sequence 6, Appli
44	216.5	4.5	685	3	US-08-872-855-2	Sequence 2, Appli
45	215	4.4	1065	2	US-08-400-159-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1

US-08-616-844-40
; Sequence 40, Application US/08616844
; Patent No. 5849578
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A.
; TITLE OF INVENTION: COMPOSITION AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/616,844
; FILING DATE: 15-MAR-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,654
; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-053
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1481 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-616-844-40

```
Query Match          98.7%; Score 4798; DB 2; Length 1481;
Best Local Similarity 90.3%; Pred. No. 0;
Matches 945; Conservative 1; Mismatches 0; Indels 100; Gaps 1;

Qy 1 MSQTETVRSVAPMRGGEITAHWLLTNTSTTSADVGTGSSASYPEGVNASVLTFQSDSTVQS 60
Db 436 MSQTETVRSVAPMRGGEITAHWLLTNTSTTSADVGTGSSASYPEGVNASVLTFQSDSTVQS 495
Qy 61 GGSHTALGDRSYSSSSSTSSSLSNAPRGERS----- 94
Db 496 GGSHTALGDRSYSSSSSTSSSLSNAPRGERSLTLEDSREPQALGDSSANAEDRTSGV 555
Qy 95 ----- 94
Db 556 PSLGTHTLATVGTNGERTLRSVLTNTSMSTTSGEAGSPAAAMPQETEGASLHVNTDDM 615
Qy 95 -----IAGISYQVGRGTAIEQRTSSDHTDHTYLSSTFTKGERALLSITDNS 140
Db 616 GLVSRSLAASSALGVAGISYQVGRGTAIEQRTSSDHTDHTYLSSTFTKGERALLSITDNS 675
Qy 141 SSSDIVESSTSYKISNSHSEYSSSFSHAQTERSNISSYDGEYAQPTESPVLHTSNLPS 200
Db 676 SSSDIVESSTSYKISNSHSEYSSSFSHAQTERSNISSYDGEYAQPTESPVLHTSNLPS 735
Qy 201 YTPITNNPNTSVVLDTDAEFVSDSSSSSSSSSSSGPPLPLPSVSQSHHLFSSILPST 260
Db 736 YTPITNNPNTSVVLDTDAEFVSDSSSSSSSSSSSGPPLPLPSVSQSHHLFSSILPST 795
Qy 261 RASVHLKLTSDASTPMSSSPSPPLVSLTSTSAPLSVSQTTLPQSSSTPVLPRARETPV 320
Db 796 RASVHLKLTSDASTPMSSSPSPPLVSLTSTSAPLSVSQTTLPQSSSTPVLPRARETPV 855
Qy 321 TSQOTSTMTSFMTHLHSQTLADLSQSTPHQEKVITESKSPSLVSLPTESTKAVTTNSPL 380
Db 856 TSQOTSTMTSFMTHLHSQTLADLSQSTPHQEKVITESKSPSLVSLPTESTKAVTTNSPL 915
Qy 381 PPSLTESSTEQTLPATSTNLNAQSPSTFTTILKTSQPLMTTPGTLSTASLVTGPIAVQT 440
Db 916 PPSLTESSTEQTLPATSTNLNAQSPSTFTTILKTSQPLMTTPGTLSTASLVTGPIAVQT 975
Qy 441 TAGQLSLTHPEILVPOISTEGGISTERNRIVVDATTLGLIPLTSVPVTSKEMTTKLGVTA 500
Db 976 TAGQLSLTHPEILVPOISTEGGISTERNRIVVDATTLGLIPLTSVPVTSKEMTTKLGVTA 1035
Qy 501 EYSPASRLGTSPPQTVVYSTAEDLAPKATFAVQSSSTQSPITLSSASVNSCAVNPCL 560
Db 1036 EYSPASRLGTSPPQTVVYSTAEDLAPKATFAVQSSSTQSPITLSSASVNSCAVNPCL 1095
Qy 561 HNGECVADNTSRGYHCRPPSWQDDCSVDVNECLSNPCPSTATCNTNQTGFSFCKCPVGY 620
Db 1096 HNGECVADNTSRGYHCRPPSWQDDCSVDVNECLSNPCPSTATCNTNQTGFSFCKCPVGY 1155
Qy 621 QLEKGINLVRTFVTEPKLKRTEFLNTFVKHSDLQVEVENEITKTLNCFPSALPSYIRSTV 680
Db 1156 QLEKGINLVRTFVTEPKLKRTEFLNTFVKHSDLQVEVENEITKTLNCFPSALPSYIRSTV 1215
Qy 681 HASRESNAVVISLQTTFSLASNTVFLDLADMOKVCNSCKSSAEVCCOLLGSRRIFRAGS 740
Db 1216 HASRESNAVVISLQTTFSLASNTVFLDLADMOKVCNSCKSSAEVCCOLLGSRRIFRAGS 1275
Qy 741 LCKRKSPECDKTSICTDLGVALCOCKSGYFQFNKMDHSCRACEDGYRLENETCMSCPF 800
Db 1276 LCKRKSPECDKTSICTDLGVALCOCKSGYFQFNKMDHSCRACEDGYRLENETCMSCPF 1335
Qy 801 GLGLNCGNPYQLITVVVIAAAGGGLLILGIALIVTCCRKNKNDISKLIKFGSGDFQMSPY 860
Db 1336 GLGLNCGNPYQLITVVVIAAAGGGLLILGIALIVTCCRKNKNDISKLIKFGSGDFQMSPY 1395
Qy 861 AEYPKPNRSEGWREATEMHENGSTKNLQMTDVIYSPTSVRNPELERNGLYPAYTCLGP 920
Db 1396 AEYPKPNRSEGWREATEMHENGSTKNLQMTDVIYSPTSVRNPELERNGLYPAYTCLGP 1455
```

```
Qy 921 SRHSCIPPGOYNPSFISDESRRDYF 946
Db 1456 SRHSCIPPGOYNPSFISDESRRDYF 1481

RESULT 2
US-08-599-654-40
; Sequence 40, Application US/08599654
; Patent No. 5882925
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/599,654
; FILING DATE: 09-FEB-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-041
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1481 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-599-654-40
```

```
Query Match          98.7%; Score 4798; DB 2; Length 1481;
Best Local Similarity 90.3%; Pred. No. 0;
Matches 945; Conservative 1; Mismatches 0; Indels 100; Gaps 1;

Qy 1 MSQTETVRSVAPMRGGEITAHWLLTNTSTTSADVGTGSSASYPEGVNASVLTFQSDSTVQS 60
Db 436 MSQTETVRSVAPMRGGEITAHWLLTNTSTTSADVGTGSSASYPEGVNASVLTFQSDSTVQS 495
Qy 61 GGSHTALGDRSYSSSSSTSSSLSNAPRGERS----- 94
Db 496 GGSHTALGDRSYSSSSSTSSSLSNAPRGERSLTLEDSREPQALGDSSANAEDRTSGV 555
Qy 95 ----- 94
Db 556 PSLGTHTLATVGTNGERTLRSVLTNTSMSTTSGEAGSPAAAMPQETEGASLHVNTDDM 615
Qy 95 -----IAGISYQVGRGTAIEQRTSSDHTDHTYLSSTFTKGERALLSITDNS 140
Db 616 GLVSRSLAASSALGVAGISYQVGRGTAIEQRTSSDHTDHTYLSSTFTKGERALLSITDNS 675
Qy 141 SSSDIVESSTSYKISNSHSEYSSSFSHAQTERSNISSYDGEYAQPTESPVLHTSNLPS 200
Db 676 SSSDIVESSTSYKISNSHSEYSSSFSHAQTERSNISSYDGEYAQPTESPVLHTSNLPS 735
Qy 201 YTPITNNPNTSVVLDTDAEFVSDSSSSSSSSSSSGPPLPLPSVSQSHHLFSSILPST 260
Db 736 YTPITNNPNTSVVLDTDAEFVSDSSSSSSSSSSSGPPLPLPSVSQSHHLFSSILPST 795
Qy 261 RASVHLKLTSDASTPMSSSPSPPLVSLTSTSAPLSVSQTTLPQSSSTPVLPRARETPV 320
Db 796 RASVHLKLTSDASTPMSSSPSPPLVSLTSTSAPLSVSQTTLPQSSSTPVLPRARETPV 855
Qy 321 TSQOTSTMTSFMTHLHSQTLADLSQSTPHQEKVITESKSPSLVSLPTESTKAVTTNSPL 380
Db 856 TSQOTSTMTSFMTHLHSQTLADLSQSTPHQEKVITESKSPSLVSLPTESTKAVTTNSPL 915
Qy 381 PPSLTESSTEQTLPATSTNLNAQSPSTFTTILKTSQPLMTTPGTLSTASLVTGPIAVQT 440
Db 916 PPSLTESSTEQTLPATSTNLNAQSPSTFTTILKTSQPLMTTPGTLSTASLVTGPIAVQT 975
Qy 441 TAGQLSLTHPEILVPOISTEGGISTERNRIVVDATTLGLIPLTSVPVTSKEMTTKLGVTA 500
Db 976 TAGQLSLTHPEILVPOISTEGGISTERNRIVVDATTLGLIPLTSVPVTSKEMTTKLGVTA 1035
Qy 501 EYSPASRLGTSPPQTVVYSTAEDLAPKATFAVQSSSTQSPITLSSASVNSCAVNPCL 560
Db 1036 EYSPASRLGTSPPQTVVYSTAEDLAPKATFAVQSSSTQSPITLSSASVNSCAVNPCL 1095
Qy 561 HNGECVADNTSRGYHCRPPSWQDDCSVDVNECLSNPCPSTATCNTNQTGFSFCKCPVGY 620
Db 1096 HNGECVADNTSRGYHCRPPSWQDDCSVDVNECLSNPCPSTATCNTNQTGFSFCKCPVGY 1155
Qy 621 QLEKGINLVRTFVTEPKLKRTEFLNTFVKHSDLQVEVENEITKTLNCFPSALPSYIRSTV 680
Db 1156 QLEKGINLVRTFVTEPKLKRTEFLNTFVKHSDLQVEVENEITKTLNCFPSALPSYIRSTV 1215
Qy 681 HASRESNAVVISLQTTFSLASNTVFLDLADMOKVCNSCKSSAEVCCOLLGSRRIFRAGS 740
Db 1216 HASRESNAVVISLQTTFSLASNTVFLDLADMOKVCNSCKSSAEVCCOLLGSRRIFRAGS 1275
Qy 741 LCKRKSPECDKTSICTDLGVALCOCKSGYFQFNKMDHSCRACEDGYRLENETCMSCPF 800
Db 1276 LCKRKSPECDKTSICTDLGVALCOCKSGYFQFNKMDHSCRACEDGYRLENETCMSCPF 1335
Qy 801 GLGLNCGNPYQLITVVVIAAAGGGLLILGIALIVTCCRKNKNDISKLIKFGSGDFQMSPY 860
Db 1336 GLGLNCGNPYQLITVVVIAAAGGGLLILGIALIVTCCRKNKNDISKLIKFGSGDFQMSPY 1395
Qy 861 AEYPKPNRSEGWREATEMHENGSTKNLQMTDVIYSPTSVRNPELERNGLYPAYTCLGP 920
Db 1396 AEYPKPNRSEGWREATEMHENGSTKNLQMTDVIYSPTSVRNPELERNGLYPAYTCLGP 1455
```


Db	616	GLVSRSLAASALGVAGISYGVQVRGTAIQORTSSDHTDHTYLSSTFTTKGERALLSITDNS	675
Qy	141	SSSDIVESSTSYIKINSNSHSEYSSFSHAQTERSNISYDGEYAQPSTESPVLHTSNLPS	200
Db	676	SSSDIVESSTSYIKINSNSHSEYSSFSHAQTERSNISYDGEYAQPSTESPVLHTSNLPS	735
Qy	201	YTPINPNMTSVVLDTDAEVPVSDSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS	260
Db	736	YTPINPNMTSVVLDTDAEVPVSDSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS	795
Qy	261	RASVHLLKSTSDASTWPSSSPPLPVSLTSTTSAPLSYSOITTLPOSSSTPVLPRARETPV	320
Db	796	RASVHLLKSTSDASTWPSSSPPLPVSLTSTTSAPLSYSOITTLPOSSSTPVLPRARETPV	855
Qy	321	TSFQTSMTSFMTMLHSSQTADLKQSOTPHQEKVITESKSPSLVSLPTSTKAVTTSNPL	380
Db	856	TSFQTSMTSFMTMLHSSQTADLKQSOTPHQEKVITESKSPSLVSLPTSTKAVTTSNPL	915
Qy	381	PPSITESSTEQTLPATSTNLQMSPTFTTTILKTSQPLMTTPTGLTSLSTASLVTGPVAVQT	440
Db	916	PPSITESSTEQTLPATSTNLQMSPTFTTTILKTSQPLMTTPTGLTSLSTASLVTGPVAVQT	975
Qy	441	TAGQKLSLTHPEILVPOISTEGGISTERNRIVDATGLPLT SVPTSAKEMTTKLGVTA	500
Db	976	TAGQKLSLTHPEILVPOISTEGGISTERNRIVDATGLPLT SVPTSAKEMTTKLGVTA	1035
Qy	501	EYSPASRLSGTSPQTTVYSTAEDLAPKSAFVQOSSTQSPSTTLLSSASVNSCAVNPL	560
Db	1036	EYSPASRLSGTSPQTTVYSTAEDLAPKSAFVQOSSTQSPSTTLLSSASVNSCAVNPL	1095
Qy	561	HNGECVADNTSRGYHCRCPPSMQGGDCSYDVNCECLSNPCPSTATCNETGGSFICKCPVGY	620
Db	1096	HNGECVADNTSRGYHCRCPPSMQGGDCSYDVNCECLSNPCPSTATCNETGGSFICKCPVGY	1155
Qy	621	OLEXGICNLVRTFVTEFKLRTFLNTVBEKHSDLQEVENEITKTLNMCFSALPSYIRSTV	680
Db	1156	OLEXGICNLVRTFVTEFKLRTFLNTVBEKHSDLQEVENEITKTLNMCFSALPSYIRSTV	1215
Qy	681	HASRESNAVVISLQTTVFLSLASNTVFLDLADRMQKCVNSCKSSAEVCQLLGSORRIFRAGS	740
Db	1216	HASRESNAVVISLQTTVFLSLASNTVFLDLADRMQKCVNSCKSSAEVCQLLGSORRIFRAGS	1275
Qy	741	LCKRKSPECDKDTSICTDLGVALCOCKSGYQFNKMDHSCRACEDGYRLENETCMSCPF	800
Db	1276	LCKRKSPECDKDTSICTDLGVALCOCKSGYQFNKMDHSCRACEDGYRLENETCMSCPF	1335
Qy	801	GLGLGNCNYPQLITVVIAAGGGLLLIILGIALIVTCCKKKNKNDISKLIFFKSGDFQMSPY	860
Db	1336	GLGLGNCNYPQLITVVIAAGGGLLLIILGIALIVTCCKKKNKNDISKLIFFKSGDFQMSPY	1395
Qy	861	AEYPKNPRSEWGREALEMHENGSTKNLQMTDVTYVPTSVRNPELRNGLYPAYTGLPG	920
Db	1396	AEYPKNPRSEWGREALEMHENGSTKNLQMTDVTYVPTSVRNPELRNGLYPAYTGLPG	1455
Qy	921	SRHSCIFPGQNPNSFISDESRRRDYF	946
Db	1456	SRHSCIFPGQNPNSFISDESRRRDYF	1481
RESULT 3			
US-08-944-868A-40			
; Sequence 40, Application US/08944868A			
; Patent No. 6018025			
; GENERAL INFORMATION:			
; APPLICANT: PALB, DEAN A			
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE			
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE			
; NUMBER OF SEQUENCES: 54			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: PENNIE & EDMONDS			
; STREET: 1155 Avenue of the Americas			
; CITY: New York			
; STATE: New York			

```

: COUNTRY: USA
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/944,868A
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/599,654
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/386,844
: FILING DATE: 10-FEB-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: CORUZZI, LAURA A
: REGISTRATION NUMBER: 30,742
: REFERENCE/DOCKET NUMBER: 7853-041
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 40:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1481 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
: US-08-944-868A-40

Query Match 98.7%; Score 4798; DB 3; Length 1481;
Best Local Similarity 90.3%; Pred. No. 0;
Matches 945; Conservative 1; Mismatches 0; Indels 100; Gaps

QY 1 MSQETVSRSVAPMRGGEITAHWLLTNTSTADVTGSSASYPEGVNASVLTFQSDSTVQS 60
DB 436 MSQETVSRSVAPMRGGEITAHWLLTNTSTADVTGSSASYPEGVNASVLTFQSDSTVQS 495
QY 61 GGSHTALGDRSYSSESSSTSSESLNSAPRGERS- ----- 94
DB 496 GGSHTALGDRSYSSESSSTSSESLNSAPRGERSTLEDSPGQALCDSSANEDRTSGV 555
QY 95 ----- 94
DB 556 PSLGTHTLATVTGNGERTLRSVTLTNTSMSTTSGEAGSPAAAMPQETEGASLHVNTDDM 615
QY 95 -----IAGISYGVQVRGTAIEORTSSDHTDHTYLLSTFTKGERALLSITDNS 140
DB 616 GLVSRSLAASGALGAVGISYGVQRGTAIEORTSSDHTDHTYLLSTFTKGERALLSITDNS 675
QY 141 SSSDIVESSTSYIKISNSHSEYSSFSHAQTERSNITSSYDGEYAQSTESPVLRTSNLPS 200
DB 676 SSSDIVESSTSYIKISNSHSEYSSFSHAQTERSNITSSYDGEYAQSTESPVLRTSNLPS 735
QY 201 YTPITNPMTSVVLDTDAEFVSDSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 260
DB 736 YTPITNPMTSVVLDTDAEFVSDSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 795
QY 261 RASVHLLKSTSDASTPWSSSPPLPVSLTWTSTNAPLSVSQTTLPQSSSTPVLPRARETPV 320
DB 796 RASVHLLKSTSDASTPWSSSPPLPVSLTWTSTNAPLSVSQTTLPQSSSTPVLPRARETPV 855
QY 321 TSFOTSTMTSPMTMLHSSQATDLKQSTPQHQKVITESKPSLVSLPTTESTKAVTNSPL 380
DB 856 TSFOTSTMTSPMTMLHSSQATDLKQSTPQHQKVITESKPSLVSLPTTESTKAVTNSPL 915
QY 381 PPSLTESSTEQTLPATSTNLAQMSPTFTTTILKTSQPLMTPTGTLSTASLVGTPIAVQT 440

```

Query Match	98.7%;	Score 4798;	DB 3;	Length 1481;
Best Local Similarity	90.3%;	Pred. No. 0;		
Matches 945;	Conservative 1;	Mismatches 0;	Indels 100;	Gaps
yy	1	MSQETVRSVAPMRGGEITAHLLLTNTTADVTGSSASYPEGVNASVLTFQFSDSTVQS	60	
bb	436	MSQETVRSVAPMRGGEITAHLLLTNTTADVTGSSASYPEGVNASVLTFQFSDSTVQS	495	
yy	61	GGSHLTALGDRSYSESSSTSSSESLNSSAPRGERS-----	94	
bb	496	GGSHLTALGDRSYSESSSTSSSESLNSSAPRGERSLTEDSREPGQALGSSANAEDRTSGV	555	
yy	95	-----	94	
bb	556	PSLGTHTLATVTCNGERTLRSVTLTNTSMKSTTSGBAGSAPAAAMPQETEGASLHVNTDDM	615	
yy	95	-----LAGISYGOVGRGTAIEQRTSSDHTDHTVLSSTFTFKGERALLSITDNS	140	
bb	616	GLVRSRLAASSALGVAGISYGOVGRGTAIEQRTSSDHTDHTVLSSTFTFKGERALLSITDNS	675	
yy	141	SSSDIVESTSYIKISNSHSHSYSSFSHAQTERNSISYDGEYAQPTESPVLHTNSLPS	200	
bb	676	SSSDIVESTSYIKISNSHSHSYSSFSHAQTERNSISYDGEYAQPTESPVLHTNSLPS	735	
yy	201	YTPTINMPNTSVVLDTDAEFVSDSSSSSSSSSSSSSGPPLPLPSVQSQSHLFSSTLPST	260	
bb	736	YTPTINMPNTSVVLDTDAEFVSDSSSSSSSSSSSSSGPPLPLPSVQSQSHLFSSTLPST	795	
yy	261	RASVHLKLTSDASTPWSSSPPLPVSLTITTSAPLSVQSTTLPQSSSTPVLPRARETPV	320	
bb	796	RASVHLKLTSDASTPWSSSPPLPVSLTITTSAPLSVQSTTLPQSSSTPVLPRARETPV	855	
yy	321	TSQTSMTSFWMLHSSQTDALKSQSTPHQEKVITESKPSLSVLTESTKAVTNTNSPL	380	
bb	856	TSQTSMTSFWMLHSSQTDALKSQSTPHQEKVITESKPSLSVLTESTKAVTNTNSPL	915	
yy	381	PPSLTESSTEQTLPATSTNLAQMSPTFTTTLKTSQPLMTTPGTSLSTASLVGTIAVQT	440	

```

RESULT      3
US-08-944-869A-40
; Sequence 40. Application US/08944868A
; Patent No. 6018025
; GENERAL INFORMATION:
; APPLICANT:  FALB, DEAN A
; TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION:  TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES:  54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  PENNIE & EDMONDS
; STREET:  1155 Avenue of the Americas
; CITY:  New York
; STATE:  New York
;

```

Db 916 PPSLTESSTEQTLPATSTNLAQSPFTTTTLKTSQPLMTTPGTLSTASLVTPGPIAVQT 975
QY 441 TAGKQLSLTHPEILVPOISTEGGISTERNRVIIVDATGTLPLTSVPTSAKEMTKLGVT 500
Db 976 TAGKQLSLTHPEILVPOISTEGGISTERNRVIIVDATGTLPLTSVPTSAKEMTKLGVT 1035
QY 501 EYSPASRLSGTSPQTTVVYSTADLAPKSAATFAVQSSQSPPTLSSASVNSCAVNPCL 560
Db 1036 EYSPASRLSGTSPQTTVVYSTADLAPKSAATFAVQSSQSPPTLSSASVNSCAVNPCL 1095
QY 561 HNGECVADNTRSGYHCRCPSPWQDDCSVDVNECLSNPCPSTATCNTQSGFICKCPVG 620
Db 1096 HNGECVADNTRSGYHCRCPSPWQDDCSVDVNECLSNPCPSTATCNTQSGFICKCPVG 1155
QY 621 QLEKGCINLVRTFTVTEFKLRTFLNTTVEKHSDLQVEVENEITKLNCFPSALPSYIRSTV 680
Db 1156 QLEKGCINLVRTFTVTEFKLRTFLNTTVEKHSDLQVEVENEITKLNCFPSALPSYIRSTV 1215
QY 681 HASRESNAVVISLQTTFSLASNTVTLFDLADRMQKCVNSCKSSAEVQQLGQRRIFRAGS 740
Db 1216 HASRESNAVVISLQTTFSLASNTVTLFDLADRMQKCVNSCKSSAEVQQLGQRRIFRAGS 1275
QY 741 LCKRKSPCKDYSICTDLGVALCCKSGYFQFNKMDHSCRACEDGYRLENETCMSCPF 800
Db 1276 LCKRKSPCKDYSICTDLGVALCCKSGYFQFNKMDHSCRACEDGYRLENETCMSCPF 1335
QY 801 GLGLNCGNPYLITVYVIAAGGGLLILGIALIVTCCRKNKNDISKLIKFGSGDFQMSPY 860
Db 1336 GLGLNCGNPYLITVYVIAAGGGLLILGIALIVTCCRKNKNDISKLIKFGSGDFQMSPY 1395
QY 861 AEPKPNRQEWGREATEMHENGSTKNLQMTDVYYSPTSVRNPPELERNGLYPAYTGLPG 920
Db 1396 AEPKPNRQEWGREATEMHENGSTKNLQMTDVYYSPTSVRNPPELERNGLYPAYTGLPG 1455
QY 921 SRHSCLFPGQYNPSFSDSERRDYF 946
Db 1456 SRHSCLFPGQYNPSFSDSERRDYF 1481

RESULT 4
US-08-944-423A-40
; Sequence 40, Application US/08944423A
; Patent No. 6020463
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: WINDOWS 95
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,423A
; FILING DATE: 06-OCT-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,654
; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: JUN-07-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995

; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1481 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-944-423A-40

Query Match 98.7%; Score 4798; DB 3; Length 1481;
Best Local Similarity 90.3%; Pred. No. 0;
Matches 945; Conservative 1; Mismatches 0; Indels 100; Gaps 1;

QY 1 MSQTEVSRVAPMRGGEITAHWLLTNTSTTADVTGSSASYPEGVNASVLTQFSDSTVQS 60
Db 436 MSQTEVSRVAPMRGGEITAHWLLTNTSTTADVTGSSASYPEGVNASVLTQFSDSTVQS 495
QY 61 GGSHTALGDRSYSESSSTSSSESLNSAPRGERS- - - - - 94
Db 496 GGSHTALGDRSYSESSSTSSSESLNSAPRGERS- - - - - 555
QY 95 - - - - - 94
Db 556 PSLGTHLTATVNGERTLRSVTLTNTSMSTTSGEAGSPAAAMPQETEGASLHVNVTDDM 615
QY 95 - - - - - 140
Db 616 GLVRSRLAASALGVAGISYGVQVGTAEQRTSSDHTDHTYLSSTFTFKGERALLSITDNS 675
QY 141 SSSDIVESSSTYIKISNSHSEYSSFSHAQTERSNTSSYDGEYAQPTESPVLHTSNLPS 200
Db 676 SSSDIVESSSTYIKISNSHSEYSSFSHAQTERSNTSSYDGEYAQPTESPVLHTSNLPS 735
QY 201 YPTINMPNTSVVLDTDAEFVSDSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 260
Db 736 YPTINMPNTSVVLDTDAEFVSDSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS 795
QY 261 RASVHLLKSTSDASTPWSPPSLPVSLTSTAPLSVSQTLTPOSSSTPVLPRARETPV 320
Db 796 RASVHLLKSTSDASTPWSPPSLPVSLTSTAPLSVSQTLTPOSSSTPVLPRARETPV 855
QY 321 TSFQTSTMTSFTMTLHSSQTDADLKQSQTPHQEKVITEKSPSLVSLPTESTKAVTTSNPL 380
Db 856 TSFQTSTMTSFTMTLHSSQTDADLKQSQTPHQEKVITEKSPSLVSLPTESTKAVTTSNPL 915
QY 381 PPSLTESSTEQTLPATSTNLAQSPFTTTTLKTSQPLMTTPGTLSTASLVTPGPIAVQT 440
Db 916 PPSLTESSTEQTLPATSTNLAQSPFTTTTLKTSQPLMTTPGTLSTASLVTPGPIAVQT 975
QY 441 TAGKQLSLTHPEILVPOISTEGGISTERNRVIIVDATGTLPLTSVPTSAKEMTKLGVT 500
Db 976 TAGKQLSLTHPEILVPOISTEGGISTERNRVIIVDATGTLPLTSVPTSAKEMTKLGVT 1035
QY 501 EYSPASRLSGTSPQTTVVYSTADLAPKSAATFAVQSSQSPPTLSSASVNSCAVNPCL 560
Db 1036 EYSPASRLSGTSPQTTVVYSTADLAPKSAATFAVQSSQSPPTLSSASVNSCAVNPCL 1095
QY 561 HNGECVADNTRSGYHCRCPSPWQDDCSVDVNECLSNPCPSTATCNTQSGFICKCPVG 620
Db 1096 HNGECVADNTRSGYHCRCPSPWQDDCSVDVNECLSNPCPSTATCNTQSGFICKCPVG 1155
QY 621 QLEKGCINLVRTFTVTEFKLRTFLNTTVEKHSDLQVEVENEITKLNCFPSALPSYIRSTV 680
Db 1156 QLEKGCINLVRTFTVTEFKLRTFLNTTVEKHSDLQVEVENEITKLNCFPSALPSYIRSTV 1215


```
Query Match          5.7%; Score 276.5; DB 6; Length 2409;
Best Local Similarity 20.0%; Pred. No. 2.8e-10;
Matches 182; Conservative 104; Mismatches 280; Indels 343; Gaps 36;

Qy 2 SQTETVSRVAPM-----RGEITPAHLLTNSTTSADVTGSSASYPEGVNASVLTQF 53
Db 1315 NRMENVAKEGVLVSQTDIFEGSGSVT-----STTLIELSDTGA--EGPTVAPLPFS 1365

Qy 54 SD-----STVQSGSHTALGDRSYSESSSTSSSESLSNSA----- 88
Db 1366 TDIGHPONQVRMAEEIQTSPOTITQDKNKNSSTAETNETTSSDFFLARAYFEMAK 1425

Qy 89 -----PRGERSIAGISGVQVRGTAIEQRTSSDHTD--HTYLSSTFTTGER 131
Db 1426 EFVTSAPKPSDLYVEPSEGS-----GEV-----DIVDSFHTSATQATQRES 1468

Qy 132 ALLSITONS-----SSSDIVESSTSYTKINSNSHSEYSSFSHAQTER-SNISSY--- 179
Db 1469 STTFVSDGSLKHPVPSPAKAVTAGPPTVSVMLPLHSEQNKSPDPTSLSNVTYSERS 1528

Qy 180 DGEY-----AOPS-----TESPVLHTSNLPSTV-- 202
Db 1529 TDGFSQDRPREFEDSTLKNRKKPTENIIIDDKEDKDLITITESTILEI--LPELTSD 1586

Qy 203 --PTINNPTSV-----LQTDABEFVSDSSSSSSSSSSSS----- 236
Db 1587 KNTIIDIDHTKPVYEDILGMQTDIDTEVPSEPHDSNDSNQVQEIYEAANVLSLITE 1646

Qy 237 ---SGPPLPLPSVSQSHLFLSSILPSTRASVHLKSTSDASTPSSPSPPLVSLTSTS 293
Db 1647 ETEFGSADVLASTQATH--DESMYEDRQLDHMGHFHTTGIPAPSTETELDVLLPTATS 1705

Qy 294 APL--SVSQTLTPOSSSTPVLPRARETEVTSFOTSTMTSFMTMLHSSOTADLKQSTPHQE 352
Db 1706 LPTRKATVIPETEGTKAKALD---DMFESTLSDGQAIADQSILITLQGFERTQE 1762

Qy 353 KVITESKPSLVSLPTE-----STKAVTNSPLPPLSFLESSTEQTLPATSNLAQMSPTFT 408
Db 1763 EY--EDKKHAGPSQPEFSSGABEALVDHTPY-----LSIATTHLMDQSVTE 1808

Qy 409 TTILKTSQPLMTPTCL--SSTASL-----VTGPVAVOTTAGKQLSLTHPEILYPOISTEG 462
Db 1809 PDVMEGNSPPYYTDTTLAVSTFAKLSSQTPSSPLTIY--SGSEAS-CHTEI--POPSALP 1863

Qy 463 GI-----STERNRIVDAT-----TGLIPLTSVPVTSKAKEMTTKL 497
Db 1864 GIDVGSSVMSPQDSFKFIHVNIETATFKPSSEYVLIHTEPPLSPDFTKLEPSEDDGKPELL 1923

Qy 498 VTAEYSP-----ASRLGTSPSPQT--TVVSTAEADL----- 526
Db 1924 EEMEASPTELIAVEGTEILODFQNKTDQVSGEAKMFPTIKTPEACTVITTADEIELEG 1983

Qy 527 -----APKSATFAVOS-----STQSPFTLSSSASVN----- 552
Db 1984 ATOMPHSTASATYGVAGVVPMLSPQTSBRPTLSSSPINPETQAAALIRGQDSITIAASE 2043

Qy 553 ----- 552
Db 2044 QQVAARILDSNDQATVNPVENTEVATPPFSLLETSTNETDFLIGNEESVEGTALYLPGP 2103

Qy 553 -SCAVNPCLHNGECVADNTSRGYHCRPPSPWQGDSCSDVYNECLSNPCPSTCANNPQGS 611
Db 2104 DRCKMNPCLNGGTCYPTETS--YVCTCVPGYSGDQCELDLDFECHSNPCRNCGATCVDGENT 2161

Qy 612 FICKCPVGY 620
Db 2162 FRCLCLPSY 2170
```

RESULT 13
US-09-165-239A-4
; Sequence 4, Application US/09165239A

```
Patent No. 6344554
; GENERAL INFORMATION:
; APPLICANT: JOHNSON, ALEXANDER
; APPLICANT: BRAUN, BURKHARD R
; TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCES FROM CANDIDA
; TITLE OF INVENTION: ALBICANS ENCODING POLYPEPTIDES ASSOCIATED WITH FILAMENTOUS
; FILE OF INVENTION: GROWTH
; FILE REFERENCE: 220022000700
; CURRENT APPLICATION NUMBER: US/09/165,239A
; CURRENT FILING DATE: 1998-10-01
; PRIOR APPLICATION NUMBER: 60/068,065
; PRIOR FILING DATE: 1997-12-18
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 750
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-165-239A-4
```

```
Query Match          5.4%; Score 263.5; DB 4; Length 750;
Best Local Similarity 22.4%; Pred. No. 4e-10;
Matches 153; Conservative 100; Mismatches 238; Indels 193; Gaps 29;

Qy 44 GYNASVLTQFSDSTVQSGGSHALGDRSYSESSSTSSSESLSNSAPRGER----SIAGIS 99
Db 80 GYNAGNAFAFVVSNAKK-----LSDGSGYGDICNFKSDSSVOLNLAFGKKVKQLSITGTG 133

Qy 100 YQVQR-----GTAIEQRTSSDHTHTYLSSTFTTGERALLS-----ITD---NSSSSDIVE 147
Db 134 YSDISLGLGNVANPPEWSAS-----LKVKAIEIVKGCCLPSGFRIVTDFESNCPEFDAIK 187

Qy 148 ----SSSYTKISNSHS---EYSSFSHAQT-----ERSNIS----- 177
Db 188 QFQSSQIIYKNAVNAIGTFDASALFNAQVKAFAKPAKRELDDEEELNSDGVTHSKRTL 247

Qy 178 -----SYDGEYAQFSTESPVYLHTSNLPSYTPITNMPNTSVVLDTDA 218
Db 248 LLLGLLKKVTGGCDTLQQFCWDCQDTPSPSTTTVTSTSSAPSTSPSSAPSTTTVTSSS 307

Qy 219 EFVSDSSSSSSSSSSSSSGP-PLPLPSVSQSHHLFSLILPSTRASVHLKSTSDASTPW 277
Db 308 PVTSPESPVPETTTVTSSVPETTPESSAPETTTVTSSVPETTPESSAPETTPESSAPE 367

Qy 278 SSSPSPLVSLTTSAPLSVSQTLTPOSSS---TPVLPRARETPTVTSFOTSTM----- 328
Db 368 SSVPESSAPETTPESSAP-----ESSVPESSAPETETETPTAHLTTTITA-OTTIVITVS 422

Qy 329 -----TSFMTMLHSSQATDLKS-----QSTPHQEKVI 355
Db 423 CSNNACSKTEVTGCVVVTSEDITYTTFCLPLTETTPVPSSVDSTSVTSAPETTP--ESTA 480

Qy 356 TESKPSPLVSLPTESTKAVTNSPLPPLSLT---ESSSEQILPATSNLAQMSPTFTTIL 412
Db 481 PESSAPE--SSAPESS--APVTETPTGPVSTVTEQSKITVITITCSNNACSESKVTGTVV 537

Qy 413 KTSQ-----PLM--TPGTLSTASILVTGPIAQVQTAGKQLSLTHPEILVPOI-----S 459
Db 538 VTSEDITYTTFCLPLTETTPATESAPESAPESAPESAPVPA---PESSAPETETAPA 594

Qy 460 TEGGISTERNRIVDA--TTGLIPLTISVPTSAKEMTTKLGVTAEYSPASRLSGTSPSPQTT 518
Db 595 TESAPATESPPVAPGTETSPVAPESSAPAT-----ESAPATESPPVAPGTETT 642

Qy 519 VVSTAEAD-----LAPKSATFAVOSSTQSP-----TTLSSSASVNSC----- 554
Db 643 PATPGAESTPVAPVAPESSAPVAPVAPGVETTPVAPVAPSTAKTALSALVSTEGTIP 702

Qy 555 -----AVNPCLHNGECVADNTS 571
Db 703 TTLESVPATQIOPSANSSYTIASVSS 726
```



```

RESULT 14
US-08-185-432-16
; Sequence 16, Application US/08185432
; Patent No. 5750652
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Busseau, Isabelle
; APPLICANT: Diederich, Robert J.
; APPLICANT: Xu, Tian
; APPLICANT: Matsuno, Kenji
; TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/185,432
; FILING DATE: 21-JAN-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2471 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-185-432-16

```

[illegible]

```

Db      1060  CQTLVNLCSPCKNCTCVQKKAESQCLCPSGWAGAYCDVPN--VSCDIAASRRGLV 1116
RESULT 15
US-08-083-590A-19
; Sequence 19, Application US/08083590A
; Patent No. 5786158
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins
; TITLE OF INVENTION: Nucleic Acids
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,590A
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2471 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; PS-08-083-590A-19

```

Query Match	5.2%	Score 254;	DB 1;	Length 2471;
Best Local Similarity	23.7%;	Pred.	No. 1e-08;	
Matches	85;	Conservative	37;	Mismatches 119; Indels 118; Gaps 14;
Qy	550	SVNSCAVPCLHNGECVADNTS-----		571
Dz	795	NIDECASNPCLNQGTCFDDISGYTCHCVLPYTGKNCQTVLAPCSNPNPCENAAVCCKESPNF		854
Qy	572	RGYHCRCPPSHQGDDCSDVDVNECLSNCPSPATCNNYQSGFICKPVGYYQ-----		621
Dz	855	ESYTCLCAPGQQQRCTIDIDECISKPMHGLCHNTQSYMCPCPGFGMDCDEEDIDD		914
Qy	622	-----LEKGTC-NLVRTF-----VTEFKLRTELNTTVKEHSDLQEVENETTKTLNMCF		670
Dz	915	CLANPCONGSCMGVNFTFCSLCLPGF-----TGDKCOTDMNEUSEPKCKNGGTCD		966
Qy	671	ALPSYIRST-----VHASRESNAVISLOTPTFSLASNVTLFDLADRMRKVNSCKSSAE		724
Dz	967	VVNSYTKCACAGPDGVHCENNINEC-----TESCFNGGT-----CVDGINSEFC		1011
Qy	725	VCQLLGSRRIFRAGSLCKRKRSPECDK-----DTSICTDLGDVALCQCKSGYQFNKM DHS		780
Dz	1012	LCP-----VGFTGSFCLHEINCESSHPCLNEGTCVDGLGTYRCSPLGYTGKN-----		1059
Qy	781	CRA-----CEDGYRLNETCMS-----CPFGLGLGNCGNPYOILITVVIAAAGGGLLL		827

Db 1060 CQTLVNLCRSRSPCKNKGTCYQKKAESQCLCPGSGWAGAYCDVPN--VSCDIAASRRGVLV 1116

Search completed: July 23, 2002, 14:29:26
Job time: 140 sec

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	409	8.4	1104	2	S59310	probable membrane
2	399	8.2	1367	1	S48478	glucan 1,4-alpha-g
3	392.5	8.1	3507	2	T34513	hypothetical prote
4	387.5	8.0	2232	2	T34434	hypothetical prote
5	364	7.5	2271	2	F90073	hypothetical prote
6	357.5	7.4	1032	2	T34433	hypothetical prote
7	347	7.1	528	2	I47141	gastric mucin (clo
8	343	7.1	1161	2	S57180	probable membrane
9	342.5	7.0	573	2	A33533	cell surface glyco
10	339	7.0	786	2	T16509	hypothetical prote
11	328	6.8	1257	2	S28764	neurocan precursor
12	325	6.7	1306	2	S23770	MSB2 protein - yea
13	323	6.6	4776	2	E95206	cell wall surface
14	315.5	6.5	1268	2	E95201	neurocan - mouse
15	313.5	6.5	534	2	T39903	serine-rich protei
16	313	6.4	1802	2	S69703	HKRI protein precu
17	312.5	6.4	1630	2	A35577	ascites stialoglyco
18	310.5	6.4	948	2	T11678	hypothetical prote
19	310.5	6.4	3570	2	T45025	mucin MUC5B, trac
20	309.5	6.4	1169	2	S38181	flocculation prote
21	309.5	6.4	3562	2	A47171	chondroitin sulfat
22	305	6.3	725	2	A41258	a-agglutinin core
23	303.5	6.2	1643	2	T14274	versican precursor
24	301	6.2	1367	2	S51959	hypothetical prote
25	300.5	6.2	1275	2	T33369	hypothetical prote
26	297	6.1	1459	2	T33271	hypothetical prote
27	295.5	6.1	610	2	PN0012	mucin 4, tracheal
28	293.5	6.0	796	2	T21460	hypothetical prote
29	293	6.0	1609	2	S25345	probable membrane


```

RESULT      4
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34434
R:Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: Z21525
A:Accession: T34434
A:Status: Preliminary; translated from GB/EMBL/DDDBJ
A:Molecule type: DNA

```

RESULT 4
T34434
hypothetical protein K06A9.1a - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 29-Oct-1999 #sequence_revision 29-oct-1999 #te
C:Accession: T34434
R:Geisel, C.; Gattung, S.
A:Description: The sequence of *C. elegans* cosmid K06A9
A:Reference number: Z21525
A:Accession: T34434
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA


```
Db 1528 S--LSDSTSASMQSSDSQSSTASLSLSDSLSTST--NRMSTIASLSLSTSVSTSESSTSES 1584
QY 493 TTKLCVT---AEYSPASRLCTSPQTTVVSTAEDLAPKATFAVOSSTQSTPTTLSSS 548
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1585 TSEDSTSTSLSDSQSTSRSTSGASSTSTSDSRSTASTSTSTSDSRSTSDSQSMLS 1644
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 549 ASVNSCAVNPCLHNGECVADNTSGYHCRCPSPSQGDDCVSDVNECLS-NPCPSTATCEN 607
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1645 TS-TSTMSDSTSLSDSVSDSTST-----DSTASSTSGMSVSLSDSTSTST 1690
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 608 TQGFICKCPGVQLEKIGNLVRTVFTEPKLRTFLNTVTEKHSDLOEVEINEITKTLNM 667
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1691 TSASEVMSASI-----SDSQSMSESVNDSESV 1717
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 668 CFSALPYSIRSTVHASRESNAVVISLOTTF---SLASNVTFLDLADMOKVCVNSCKSSA 723
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1718 SESNESDSKMSGSTSVSDGSLUSVSTSLRKSESVSESSLSGQSMDSVSTSDSSSL 1777
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 724 EVCOLLGQRRIFRAGSLCKRCKSPCEDKDTST 754
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1778 SVSTSLASSESVSESDSLSDSKSTSGSTSTST 1808
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 6
T34433
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34433
R:Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: 221525
A:Accession: T34433
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1032 <GEI>
A:Cross-references: EMBL:U08046; PIDN:AAC70889.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 839/1; 849/1; 875/3; 940/2; 966/2; 99

Query Match 7.4%; Score 357.5; DB 2; Length 1032;
Best Local Similarity 22.4%; Pred. No. 2.9e-09;
Matches 207; Conservative 143; Mismatches 352; Indels 221; Gaps 34;

QY 6 TVSRSVAPMRGGEITAHWLLTNSTT---SADVTGSSASYPEGVNANSLVLTQFSDSTVQS 60
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 246 SITSALPIASSASSPSAASSTTPVVLSSSTIQSSSGTFPSSVASSPSTVGTSGAAS 305
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 61 GGSHTALGDRSISSESTSTSS---SSESLNSAPRGERSIAGISYGVQVRGTAIEORTSDH 116
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 306 SSVATVSTIAGTSGTITPVGSSSTIGSTSPASSSSSGTM-----STISGSGSTV 359
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 117 TDHTYLSSTTKGRALLSTDNSSDIVESST--SYIKISNSHSEYSSFSHAQTERS 174
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 360 TVVPGSSSTF-----ASSTPIASSSSPGSTVTVAPGSSSTYGSSTPSASSSS 406
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 175 N-ISSYDGEYAQPSTESPVLHSTNLSPTTINMPTSVVLDTAEFVSDSSSSSSSS 232
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 407 SGMNSTNGSTGTSTVTVAPV-SSSTFGSTPTIASSSSSG---STVTVVSGSSSYGSGST 461
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 233 SSSSSGPPPLPSPVQSQHLLFSSILPSTRASVHLKSTSDASTPSSSPSPPLVSLTST 292
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 462 PSASSSSAGTASTISGSGTATIVPGSSSV-----GSSTQASPS-PGTMT-V 511
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 293 SAPLSVSQTTLPQSSSTPV---LPRARETPVTSQTSTMTSEMTMLRSSQTADLKSGSTP 349
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 512 SGPTGSTVTVVPGSSTSPAPSSPNPSSSPASTGSTITISGSSSIIIVSTVSGSTVSGSTG 571
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```

```
QY 350 HOEKVITESKSP--SLYSLPTESTKAVTNSPLP--PSLTESSTEQTLPATSTNLAQMSP 405
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 572 TSQSTLASSATPQSSSTVPSSSSQPSPAPNTGTTTPSQTHSSQSPSPSMMPSSSTP 631
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 406 TFFTTLIKTSOPLMTTPTGLSTASLVGT---PIAVQTTAGKQLSLTHPEILVPIQISTEG 462
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 632 T-----GSSQSTITPEGSTASSPTGSGTSTFVATEVTS-----QSTVPSGSSSLG 676
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 463 GISTERN---RVIVDATTGLIPLTSVTSKEMTKLGVTAETSPASRSLSGTSPPQTTV 519
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 677 TQSTNSPSPSSLSPTSGMSTLTSEPS-----SSTQSSGAQSTLTTPSPN--- 723
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 520 VSTAEDLAPKATFAVOSSTQSTPTTLSSASVNSCAVNPCLHNGECVADNTSRGYHCRCP 579
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 724 -----PSQSTSSLESSTSGATTSSGSAGTTMTS----- 751
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 580 PSMWGDDCVSDVNECLSNPCPSTATCENNT-OGSPICKCPGVQLEKICNLVTRTFVTEFK 638
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 752 PS---QSSSVGSSQGSTSPAASSTTSGEMTSQGS-----TQTPGSSSVS 790
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 639 LKRTFLNTVTEKHSDLOEVEINEITKTLNMFCSALPSYIRSTVHASRESNAVVI----- 691
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 791 TSAAILTSTQGSVS-TNSPGSTVTRP-----STVSGSTSGSTVTVGSTEAS 836
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 692 ----SLOTFTSLASNVTFLDLADMOK-----CVNSCKSSAEVCQLLGSORRI----- 735
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 837 TSGSSLATTSAPKPSVTCFLPMWDTQSKIEDOTAITNTYTFENFALLVASKLNNEISILTY 896
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 736 ----FRAGSLCKRCKSPEC-----KDTSICTDLGDVALCQCKSGYQFQNMK 777
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 897 IDNFGYSAGLNDHQYPTDDYNGIKSVFPIDGTDGDDDLDKDV----- 941
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 778 DHSCRACEDGYRLENETCM-----SCPRG-IGGLNCGHPYQLITVVIAAAGGLLILGIA 832
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 942 DKSLATADWTPPVADQTCMIFISAPEDEYGGTTIKSTTYTFETVVGVLVGGAKSIPGLS 1001
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 833 ----LIVTCCRKNKNDISKLFK 851
| : : | : | : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 1002 IDKNIVITNTMNDRDASAVVSK 1024
| : : | : | : | : : | : : | : : | : : | : : | : : | : : | : : |

RESULT 7
I47141
gastric mucin (clone PGM-2A) - pig (fragment)
C:Species: Sus scrofa domestica (domestic pig)
C>Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 03-Nov-2000
C:Accession: I47141; S55315
R:Turner, B.S.; Bhaskar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.
Gastroenterology 106, 200, 1994
A:Title: Pig gastric mucin: isolation and characterization of a cDNA clone with a nov
A:Reference number: I47141; MUID:94102478
A:Accession: I47141
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-528 <TUR>
A:Cross-references: EMBL:U10281; NID:g915205; PIDN:AAC48526.1; PID:g915208
R:Turner, B.S.; Bhaskar, K.R.; Hadzopoulou-Cladaras, M.; Specian, R.D.; LaMont, J.T.
Biochem. J. 308, 89-96, 1995
A:Title: Isolation and characterization of cDNA clones encoding pig gastric mucin.
A:Reference number: S55315; MUID:95275264
A:Accession: S55315
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-528 <TU2>
A:Cross-references: GB:U10281; NID:g915205; PIDN:AAC48526.1; PID:g915208
C:Superfamily: pig submaxillary mucin

Query Match 7.1%; Score 347; DB 2; Length 528;
Best Local Similarity 27.9%; Pred. No. 4e-09;
Matches 161; Conservative 81; Mismatches 204; Indels 132; Gaps 23;

QY 10 SVAPMRGGEITAHWLLTNSTTSADVTGSSASYPEGVNANSLVLTQFSDS-----TVQSGG 62
```

```
Db 35 SVQPSSSGAP-----TTSATVQ--TSSSSPPISSITSVQSSSSVPTTSTTSVQPS 88
Qy 63 SHTALGDRSYSESSSTSSSLNSSAPRGERSIAGISGQVGRTAIEQRTSSDHTHYL 122
Db 89 SSSAPITRATSVQSSSSSAPISSTT-----SVQPSGGSVPTTSATSVQSSSSSAPT 143
Qy 123 SSTFTKGERALLSTIDNSSDDIVESSTSYKISNSHSEYSSFSHAQTERSNISSYDGE 182
Db 144 SAT-----SVQPSSSSPPISSIVQPSSSSSAPTTSATS----- 179
Qy 183 YAOPTES--PVLHNSLNPYTPINMPNTSVVLDTDAEFDVSSSSSS-----SSSSSSS 236
Db 180 -VQPSSSSSPPI-----SSTVSQVQSSSSSVPTTTSVQPSSSSVPTTSATSVRSS 231
Qy 237 SGPPPLPLPSVQSQHLHLLSSLPSTRASVHLLKSTSDASTWSSSPPLP-----VSLTST 292
Db 232 SSSSTPIPT-----TSVQPSSSSAPTTSATS--VQPSSSSSTPIPTTSVQPSSSS 282
Qy 293 SAPLSVQTTLPQSSSTPVLPRARETPTVTFOTSTMTSMTMLHSSQTADLKOSTPHQE 352
Db 283 SAPTTSATSVQPSSSSP-----PISS-----TISVQPSSSSSSP 317
Qy 353 KVITESKPSLV-SLPTESTKAVTNSPLPPLPSTESTEOTLPATSNLAQMSPTFTTTI 411
Db 318 TTSTTSVQPSSSGAPTTSATSVQPSSSSPPI--SSTISVQPSSSSS--SPTTSTTS 371
Qy 412 LKTSOPLMTPTGTLSSPASVLVGTGPIAVQTAGKQLSLTHPEILVPOQLSTEGGISTERNRV 471
Db 372 VQPS-----SSGAPTTSATSVQPSSSSS-----VPTTSATSVRSSSSST 412
Qy 472 IVDATGCLIP--LTSVPT-----SAKEMTKLGVTAEYSPASRSGLTSPSPQTTVVST 522
Db 413 PIPTTSVQPSSSSVPTTSATSVQSSSSTPIPTTSVQPSSSSAPTTS-ATSVQPS 471
Qy 523 AEDLAPKSAFAVQ--SSTQSPITLSSASVNSCAVP 558
Db 472 SSSSPPISSITSVQPSSSSSSPTTTSVQPSSSGSAP 509

RESULT 8
S57180
probable membrane protein YJR151c - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein J2233; serine/threonine-rich protein YJR151c
C;Species: Saccharomyces cerevisiae
C;Date: 23-Aug-1995 #sequence_revision 08-Sep-1995 #text_change 05-Nov-1999
C;Accession: S57180
R;Scarcez, T.
submitted to the Protein Sequence Database, September 1995
A;Reference number: S57169
A;Accession: S57180
A;Molecule type: DNA
A;Residues: 1-1161 <SCA>
A;Cross-references: EMBL:Z49651; NID:g1015902; PID:g1015903; GSPDB:GN00010; MIPS:YJR151c
C;Genetics:
A;Gene: MIPS:YJR151c
A;Map position: 10R
C;Keywords: transmembrane protein
```

```
Query Match 7.1%; Score 343; DB 2; Length 1161;
Best Local Similarity 21.3%; Pred. No. 1.6e-08;
Matches 204; Conservative 147; Mismatches 392; Indels 216; Gaps 34;

Qy 3 QTEVTRSVAP-----MRGGEIT-----AHWLLNSITTSADVTGSSASYPE 43
Db 60 KTEYTPSEIAAAVFDYGDFTRTLGTISGDEVTRMTGVPMYSRLKPAI-----SSALSUD 115
Qy 44 GVNASVLQFSDSTVQSGGSHLALGD-RSYSESSSSSLNSSAPRGERSIAGISYQG 102
Db 116 GIYTAIPTSTTTTKSSTTPTTTTSTSTSTPTTSTSTSTPTTSTSTPTTSTT 175
Qy 103 VRGTAIEQRTSSDHTDHT-----YLSSTFTKGERALLSIDNSSSDIVESSSYTKISN 157
```

```
Db 176 TSTPTTSTTSTPTTSTTPTTSTTPTTSTTPTTSTTPTTSTTPTTSTTPTTSTTPTTST 235
Qy 158 SSHSEYSSFSHAQTERSNISYDGEYAQPSTE----SPVLHNSLNPSTPTPIN-MPNTSV 212
Db 236 TSTTSQSTKSTPTTSTSTSTPTTSTTPTTSTTSTPTTSTTPTTSTTPTTSTTPTTSTT 295
Qy 213 VLDTDAEFVSDSSSSSS-----SSSSSSSGPPLPLPSVQSQHLHLLSSILPSTRASVHLL 267
Db 296 ---TSSFTSSASASSSVISSTATTSTTFASLITPATSTASTDHTTSSV--STNAFTTS 350
Qy 268 KSTSDASTPMSSSPPLPVLSTLSTTSAPLSVQTT--LPQSSSTPVLPRARETPTVTFOT 325
Db 351 ATTTTSDTVISSSSP--SQVTSSAEPTTVSEVTSVVEPTRSSQVTSAAEPTTVSEFTSS 407
Qy 326 STMTSFMTHLSS-----QTADLKSOQTPHQEKVITEKSPSLVS-----LPTESTKAVT 375
Db 408 SVEPTRSSQVTSAAEPTTVSEFTSSVEPTRSSQVTSAAEPTTVSEFTSSVEPTRSSQVTS 467
Qy 376 TNSPLPPLSTESSTEOTLPATSTNLAQMSPTFTTTILKTSQPLMTPTGTLs---STASLV 432
Db 468 SAEPTTVSEFTSSVEPTRSSQVTSAAE--PTTVSEFTSSVEPIRSSQVTSAAEPTTVSEV 525
Qy 433 TGPVAVOTTACKQLUSLTHP-----EILVPOQLSTEGGISTERNRVIVDA-- 475
Db 526 TS--SVEPIRSSQVTTTTEPVSSFGSTFSEITSSAEPLSFASKATTSAESISSNQITISSEL 583
Qy 476 TTGLIPLTSPVPTSAKEMTKLGVTAEYSPASRSGLTSPSPQTTVWSTAEADLAPKSAFA 534
Db 584 IVSVITSSSEIPSEIEVLTSGLSSSVVEPTSL---VGPSSDESISSTESLATSATFTSA 640
Qy 535 VQSS-----TQSPITLSSASVNSCAVN-----PCLHNGECVADNTSRGYHCRC 578
Db 641 VWSSSKAADFTTRSTVSAKSDVSCNSSTQSTTFEATPSTPLAVSSTVVTSST----- 692
Qy 579 PPSWQGGDDCVDV--NECLSNPCPSTATCNTNQCSFTCKCPVGQLEKGICNLVRT---F 633
Db 693 -----DSVSPNIPFSEISSPESSTA-ITSTSTSTAE-----RTSLSY 730
Qy 634 VTEPKLRTFLNTTVEKHSDLOEVENETKTLNMCFSALPSYI--RSTVHASRESNAVVI 691
Db 731 LSSNMSSEFTLSTFTVSQSVSSFSMEPTSSVASFSSPLLVSRRNCSDARSNTISS 790
Qy 692 SLQTTFLASLNVLT-LFDLADRMOKVNSCKSSAEVCOGLGSSQRIFRAGSLCKRKSPEC 750
Db 791 GLRFTIENVRNATSTFTNLSTDEIVITSCSS-----CT 824
Qy 751 KDTISCTDLGVALCQCKSGYFQFNKMDHSCRACEDGY----- 788
Db 825 NEDSVLTKT-----QVSTVETTITSCSGGICTTLMSPVTTINAKANTLTTE 871
Qy 789 -RLENETCMSCPGGLGGLNGPNPYQLITVIAAAGGGLLLILGIALIVTCCKRKNKNDIS 846
Db 872 TSTVETTITTCPGGV-----CSTLTVPVTTITSEA-----TTTATISCDEEDIT 917

RESULT 9
A33533
cell surface glycoprotein precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 29-Jan-1990 #sequence_revision 29-Jan-1990 #text_change 11-Jan-2000
C;Accession: A33533
R;Dougherty, G.J.; Kay, R.J.; Humphries, R.K.
J. Biol. Chem. 264, 6509-6514, 1989
A;Title: Molecular cloning of 114/A10, a cell surface antigen containing highly conse
lines.
A;Reference number: A33533; MUID:89197960
A;Accession: A33533
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-573 <DOU>
A;Cross-references: GB:J04634; NID:g191943; PIDN:AAA37239.1; PID:g309106
C;Superfamily: unassigned EGF-related proteins; EGF homology
```


A:Residues: 1-1257 <RAW>
 A:Cross-references: EMBL:M97161; NID:g205649; PIDN:AAC37679.1; PID:g205650
 C:Superfamily: aggrcan; C-type lectin homology; complement factor H repeat homology; EQ
 C:Keywords: chondroitin sulfate proteoglycan; glycoprotein
 F:1-22/Domain: signal sequence #status predicted <SIG>
 F:23-1257/Product: neurocan #status predicted <MAT>
 F:176-253/Domain: link protein repeat homology <LNK1>
 F:274-355/Domain: link protein repeat homology <LNK2>
 F:364-366/Region: cell attachment (R-G-D) motif
 F:953-984/Domain: EGF homology <EGF>
 F:1029-1149/Domain: C-type lectin homology <LCH>
 F:1156-1212/Domain: complement factor H repeat homology <FHD>
 F:121339/737,967,1164/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:372,410/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted
 F:944/Binding site: chondroitin sulfate (Ser) (covalent) #status experimental

Query Match 6.8%; Score 328; DB 2; Length 1257;
 Best Local Similarity 22.0%; Pred. No. 8.6e-08;
 Matches 155; Conservative 107; Mismatches 255; Indels 186; Gaps 31;

QY 1 MSQETVRSVAPMRGGEITAHMLLTNSTTSADVTGSSA-----SYPEGVNASVLTFQF 53
 DB 418 LTRTQASQETLASTPGPTLASMLLTGVTSSTGVPSPSSLCVDMETTPSG-----TQV 471

QY 54 SDTVOGGGSHHTALGDRSYSESSSTSSSESLNAPSAPRGERSIAGISYGQVRGTAIEQRTS 113
 DB 472 APTPTMRGRKGLNGRHFFQOQ--PEDQLLEAAEASQAQPTTEVTADHMGPSAATEALE 529

QY 114 SDHTDHTYLSSTFKGERALLSTDNSSSDIVESSTSYIKINSNHSSEYSSFSHAQTER 173
 DB 530 SDQSHSPW-----AII-----TNEVDVPCAGS-----LGSRLPESSKRWSPSLISP 570

QY 174 SNISSYD-----GEVAQSPSTESPVLTHTSNLPSYPTINMPNTSVVLDTDAEFVSDSS 226
 DB 571 STVPSTDTSTPLKPGDAEPKVASATHPPWLPFS-EPAV-----PSSIPSEALS 618

QY 227 SSSSSSSSSSSSGPLPLPSYSQSHILF-----SSILPSTRASVHLKSTSDASTPWSSSPS 282
 DB 619 AVSLQASPGDGPDPFIVAMLRAPKLWLLPHSTLVPNV-----S 657

QY 283 PLPSVLTSTTSAPLSVSQTLTPOSSSTPVLPRARETPTVTSFQSTWTFMTLHSSQATD 342
 DB 658 PIPLS-----PASPLPSSVPEEQAVRPVYSGAEDPETPQOT-TMAAPGEASHGSPED 709

QY 343 -----LKQSQTPHOEKVITESKPSLSVLPTESS--TKAVTTNSPLPPSLTESSTEQ-- 391
 DB 710 SIEIGISSHQATKHP-----ISGPWASLDSSNVTYN-PVPSDAGILGTESGV 756

QY 392 -TLPATSTNLQAQSPFTTTTILKTSQPLMTTPTGLSTASLVGTGPIAVQTAGKQLSLTH 450
 DB 757 LDLPGSPSTDGQAT---VDMVLATWLP---PCHGLDTGSQST-PMEAH---GVTKSVE- 805

QY 451 PELVLPQISTEGIGISTERNRVIIVDATTGLIPTSVTSAKE-----WTKLGVTA---EY 502
 DB 806 -----PTVALEGATKDP-----NEATMDVVPSTVDATSGSEPKSSSTSHVVVTAAGDQ 856

QY 503 SP-----ASRLGTSPTS-----PQTTVVSTAEDLAPKSAFVQ-----SSTQSP 542
 DB 857 TPTLTTSSEGVVQVESGLTSLSPHPWSSSLASDEWASVSSGEPFLWDIPSTLIP 916

QY 543 TLLSSAS-----VNSCAVNPCLHNGECVADNTRSGYHCR 577
 DB 917 VSLGLDESOLKVVAESPGLEGFEVAVASGOEDPTDPCENNPCLHGTCRTNGTM--YGCS 974

QY 578 CPSPWGGDDCSVDVNECLSNPCPSTATCNTQGSFICKCPGVY 620
 DB 975 CDQGYAGENCEIDIDCLCSPCENGCTCIDEVNGFICLCLPSY 1017

RESULT 12
 S25370
 MSB2 protein - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein G4017; protein YGR014W
 C:Species: Saccharomyces cerevisiae
 C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 21-Jul-2000
 C:Accession: S25370; S64305
 R:Bender, A.; Pringle, J.R.
 Yeast 8, 315-323, 1992
 A:Title: A Ser/Thr-rich multicopy suppressor of a cdc24 bud emergence defect.
 A:Reference number: S25370; MUID:92383951
 A:Accession: S25370
 A:Molecule type: DNA
 A:Residues: 1-1306 <BEN>
 A:Cross-references: GB:M77354; NID:g171993; PIDN:AAA34798.1; PID:g171994
 R:Rieger, M.; Mueller-Auer, S.; Brueckner, M.; Schaefer, M.
 submitted to the Protein Sequence Database, May 1996
 A:Reference number: S64071
 A:Accession: S64305
 A:Molecule type: DNA
 A:Residues: 1-1306 <RIE>
 A:Cross-references: EMBL:272799; NID:g1322977; PID:g1322978; MIPS:YGR014W
 A:Experimental source: strain S288C
 C:Genetics:
 A:Gene: SGD:MSB2
 A:Cross-references: SGD:S0003246; MIPS:YGR014W
 A:Map position: 7R
 C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida
 C:Keywords: transmembrane protein
 F:3-19/Domain: transmembrane #status predicted <TM1>
 F:1189-1205/Domain: transmembrane #status predicted <TM2>

Query Match 6.7%; Score 325; DB 2; Length 1306;
 Best Local Similarity 27.2%; Pred. No. 1.2e-07;
 Matches 151; Conservative 95; Mismatches 221; Indels 88; Gaps 21;

QY 28 STTSADVTGSSASYPEGVNASVLTFQSDSTVQSGSH-TALGDRSYSESSSTSSSESLNS 86
 DB 494 SQSSSDVATSA--PSVWSSS---FSYTLQAGGSMTPSSSTIYVSSSTGSGS--ES 545

QY 87 SAPRGERSIAGISYGQVRGTAIEQRTSSDHTHTLSSTFKGERALLSTDNSSSSDIV 146
 DB 546 AASTASATLSGSSSTYMAGNLQOPPSTS-----SLLSEQATSTSAVL 589

QY 147 ESTSYIKINSNHSSEYSSFSHAQTERSN-ISSYDGYAQP--STESPVLHTSNLPSYTP 203
 DB 590 ASS-----SVSTTSPYTTAGGASTEASSLSSTSAETSQVSVSQSTALQTSFASST 643

QY 204 TINMPNTSVVLDTDAEFVSDSSSSSSSSSSSSSGPLPLPSVYSQSHHLPSSILPSTRAS 263
 DB 644 TEGSETSSQGFSTSSVVLQMPSSLSSEFSPQTTQMNASASSSQ-----YTISS 693

QY 264 VHLKSTSDASTPWSSSPPL-----PVSLTSTAPLSVSQTLTLPQSSSTPVLPRARE 317
 DB 694 TGLSQVSDTTSVSYTTSSSSVSQVSDTPVSYTTSSSSVSQVSDTPVSYTTSSSSVSQVSD 753

QY 318 TPTVTSQTS-----TMTSEFMTLHS-SQTADLKSQSTPHQEKVITESKPSLSVLP 367
 DB 754 TPV-SYTTSSSSVSQVSDTPVSYTTSSSSVSQVSDTSPVSTSSRSSVSQVSDTP-----VP 808

QY 368 TESTKAVT--TNSPLPPLSTES-----STEQTLPATSTNLQAQSPFTTTILKTS---QP 417
 DB 809 STSRSSVSQTSLSLQPTTTSSQRTTSTHGAL-SESSSVSQQASEITSSINATASYHS 867

QY 418 LMTTPGTLST-----ASLVTPGPIAVQTAGKQLSLTHPEILVPOISTEGGISTERN- 469
 DB 868 IQTTAATQSTTLSTFDANSSASAPLEAVATSTTPSSKASLLLTPTSSLSQVATNV 927

QY 470 --RVIVDATTGLIPTSVPTSAKEMTKLGVTAEYSPASRSRLGTSQSPQTVVSTAEDLA 527
 DB 928 QTSLLTTESTVLEPSTNNSSTFSLVT--SSDNMMWIPTELITQAPAEASTASSTVGGTQ 985

QY 528 PKSATFAVQSSTQSP 542
 DB 986 TMTLPHAIATAQVP 1000

RESULT 13

E95206
cell wall surface anchor family protein [imported] - Streptococcus pneumoniae (strain T1
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: E95206
R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf, E.
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; PMID:21357209; PMID:11463916
A:Accession: E95206
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-4776 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75846.1; PID:g14973269; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP1772

Query Match 6.6%; Score 323; DB 2; Length 4776;
Best Local Similarity 24.0%; Pred. No. 7.2e-07;
Matches 137; Conservative 108; Mismatches 285; Indels 40; Gaps 9;
QY 1 MSOTETVRSVAPMRGGEITAHLWLLTNTSTTSADVGTSSASYPGVNASVLTQFSDSVQS 60
DB 2691 ISASEASTSASASASTSASASTSASASTSASASTSASASTSASASTSASESTAS 2747
QY 61 GGSHALGDRSYBESSSTSSSELNSAPRGERSIAGISYQVGRGTAIEORTSDSDHDT 120
DB 2748 STSTASASTSASESTASASTSASASTSASASTSASASTSASASTSASESTAS 2803
QY 121 YLSTFTKGERALLSITDSSSDIVESSTSYKINSNSHSEYSSFSHAOTERSNISSYD 180
DB 2804 SASASASTSASASTSASASTSASASTSASASTSASASTSASASTSASASTSASAST 2859
QY 181 GEYAPQSTSPVLHTSNLPSYTPPINPNTSVVLTDAEFVSDSSSSSSSSSSSSSSGPP 240
DB 2860 SASASASTSASESTASASTSASESTASASTSASESTASASTSASESTASASTSASA 2909
QY 241 LPLPSVQSHHLLSSILPSTRASVHLKLTSDASTPWSSSPPLPVLSTSTGAPLSVQ 300
DB 2910 SASTSASESTASASTSASASTSASASTSASESTASASTSASASTSASASTSASAS 2968
QY 301 TTLPOSSSTPVLPRARETPVTSTQTSMTSMFLH-----SSQADLKSSQSTPHQ 351
DB 2969 TSASEASTSASASTSASESTASASTSASESTASASTSASESTASASTSASESTAS 3028
QY 352 EKVITE-SKSPSLVSLPSTESKA-----VTNPSLPPLPSTESSTEQTLPLATSNLAQMS 405
DB 3029 ASASTSASASIASASESTASASESTASASTSASTSASESTASASTSASESTASASAS 3088
QY 406 TPTTTLKTSOPLMTPTGTLSSSTASVLTGPIAVQTTAGKLSLTHPILPQISTEGGIS 465
DB 3089 TSASASTSASTSASESTASASTSASASTSASASTSASASTSASASTSASASTSASV 3148
QY 466 TERNRVIVDATTGLPLTSVPTSAKEMTKLGVTAESVSPASRLGTSPTQTTVVSTAE 525
DB 3149 ASTS---ASASTSASTSASESTASASESTASASTSASESTASASTSASESTASASA 3205
QY 526 LAPKATFAVQSTQSPPTLLSSASVNSCA 555
DB 3206 STSASASTSASESTASASTSASASTSASA 3235

RESULT 14

S52781
neurocan - mouse

C:Species: Mus musculus (house mouse)
C:Date: 19-May-1995 #sequence_revision 21-Jul-1995 #text_change 04-Feb-2000
C:Accession: S52781
R:Rauch, U.; Forsberg, N.; Kulbe, G.; Arnold-Ammer, I.; Faessler, R.
submitted to the EMBL Data Library, February 1995
A:Description: Amino acid sequence of mouse neurocan and brevicin and their different
A:Reference number: S52781
A:Accession: S52781
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1268 <RAU>
A:Cross-references: PIDN:CAA59216.1; PID:g758630
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology;
F:176-253/Domain: link protein repeat homology <LNK1>
F:274-355/Domain: link protein repeat homology <LNK2>
F:964-995/Domain: EGF homology <EGF>
F:1040-1160/Domain: C-type lectin homology <LCH>
F:1167-1223/Domain: complement factor H repeat homology <FHD>

Query Match 6.5%; Score 315.5; DB 2; Length 1268;
Best Local Similarity 23.0%; Pred. No. 3.3e-07;
Matches 164; Conservative 95; Mismatches 259; Indels 195; Gaps 31;
QY 1 MSOTETVRSVAPMRGGEITA-----HWLLTNTSTTSADVGTSS-----ASYPEGVNA 47
DB 418 LTWTQAEETLTGTPGPTTLASWPSSSEKWLFTGAPSMGVSSPSDMGVDMETPLG--- 474
QY 48 SVLTQFSDSTVQSGGSHALGDRSYBESSSTSSSELNSAPRGERSIAGISYQVGRGTA 107
DB 475 ---TOVAPTFTMRGRFKGLNGRHFQOQGPEDQLPEV--AEPSAQPTTLGATANHMRPSA 529
QY 108 IEORTSSDHDTHYLSFTFKGERALLSITDSSSDIVESSTSYKINSNSHSEYSSFS 167
DB 530 ATEASESDQSHSPWAILTNEVDEFGAGSLGRSLPESLMSPS---LISPSVPS----- 580
QY 168 HAOTERNISSYDGEYAQAPSTESPVLHTSNLPSYTPPINPNTSVVLTDAEFVSDSSSS 227
DB 581 ---TE-STPKFGAARAPSVKSAIPLPLPPEPPA---PSPG-----PSEALSA 624
QY 228 SSSSSSSSSGPPPLPLPSVQSHHLLF-----SSILPSTRASVHLKLTSDASTPWSSSP-S 282
DB 625 VSLQASSADGSPDFPIVAMLRAPKLWLLPRSTLVN-----MTPLVPLSPAS 670
QY 283 PL-----PVSL-----TTSTAPLSVSQTTLPQSSSTPV---LPRARET 318
DB 671 PLPSWVPEEQAVRPSVSLGAEDLETPFTTAAAPVEASHRS-PDADSTEIEGTSSMRATKH 729
QY 319 PVTSTFQSTWTSFMTM-----LHSSQTA----- 341
DB 730 PISGPWASLDSSNVMTMPVPDAGILGTESGVLDLPGSPSGGQATVEKVLATWLPPLGQ 789
QY 342 --DLKSQSTPHQEKVITEKSPSLVSL-----PTSTKAVTNTSNPLPSPSLTESSTEQ 392
DB 790 GLDPGQSTPWEAGHVAVSMPT-VALEGGATEGPMETREV-----PSTADATWE-- 840
QY 393 LPATSTNLAQMSPTFTTILKTSOPLMTPTGTLSSSTASVLTGPIAVQTTAGKLSL-THP 451
DB 841 ----SESRSAISSTHIAVTMARAGCMPTLTSTSEGHPEPKQMVQAESLEPLNTLPSPH 896
QY 452 --ETLVPQISTEGGISTERNRVIVDATTGL--IPLTSVPTSAKEMTKLGVTAESVSPAS 507
DB 897 WSSLVVPDMEVASVSSE-----PTGLWDIPSTLIPVSLGLDSEVLNVAAE----- 942
QY 508 SLGTSFSPQTTVVSTAEADLAPKATFAVQSTQSPPTLLSSASVNSCAVNPCLHNGECVA 567
DB 943 ----SPS----VEGFWEVA-----SQEDPT-----DPCENNPCLHGTCHT 977
QY 568 DNTSRGYHCRPPSWQGDGDCSDVNECLSNPCPSTATFCNTNTQGSFICKCPVGY 620
DB 978 NGTV--YGCSCDQGYAGENCEIDIDCLCSPCENGCTCIDEVNGFICLCLPSY 1028

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 14:28:52 ; Search time 15.08 Seconds

(without alignments)
2428.959 Million cell updates/sec

Title: US-09-840-746-1

Perfect score: 4859

Sequence: 1 MSQTETSKSVAPMRGGEIT.....PPQYNPFSFISDESRDRDYF 946

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	399	8.2	1367	1 AMYH_YEAST	P08640 saccharomyc
2	396	8.1	1140	1 YMG6_YEAST	Q04893 saccharomyc
3	343	7.1	1161	1 DAN4_YEAST	P47179 saccharomyc
4	342.5	7.0	573	1 C114_MOUSE	P19467 mus musculu
5	328	6.8	1257	1 PGCN_RAT	P55067 rattus norv
6	325	6.7	1306	1 MSB2_YEAST	P32334 saccharomyc
7	315.5	6.5	1288	1 PGCN_MOUSE	P55066 mus musculu
8	315	6.5	1802	1 HKR1_YEAST	P41809 saccharomyc
9	309.5	6.4	1169	1 YK82_YEAST	P36170 saccharomyc
10	309.5	6.4	3562	1 PGCV_CHICK	Q09053 gallus gall
11	305	6.3	725	1 AGAL_YEAST	P32323 saccharomyc
12	295.5	6.1	610	1 MUC4_HUMAN	Q99102 homo sapien
13	293.5	6.0	796	1 YS8A_CAEEL	Q09625 caenorhabdi
14	293	6.0	1609	1 FIG2_YEAST	P25653 saccharomyc
15	293	6.0	3381	1 PGCV_BOVIN	P81382 bos taurus
16	291.5	6.0	797	1 VGLX_HSVEB	P28968 equine herp
17	285	5.9	1419	1 ALAL_CANAL	Q13368 candida alb
18	284.5	5.9	1537	1 FLO1_YEAST	P32768 saccharomyc
19	282.5	5.8	1322	1 YAG3_YEAST	P39712 saccharomyc
20	281.5	5.8	1260	1 ALS1_CANAL	P46590 candida alb
21	279	5.7	2524	1 NOTC_XENLA	P21783 xenopus lae
22	277	5.7	5179	1 MUC2_HUMAN	Q02817 homo sapien
23	276.5	5.7	3396	1 PGCV_HUMAN	P13611 homo sapien
24	273	5.6	1251	1 YQ03_CAEEL	Q09550 caenorhabdi
25	272.5	5.6	1075	1 FLO5_YEAST	P38894 saccharomyc
26	272	5.6	3358	1 PGCV_MOUSE	Q62059 mus musculu
27	269	5.5	1041	1 EGT2_YEAST	P42835 saccharomyc
28	267.5	5.5	2738	1 PGCV_RAT	Q9erb4 rattus norv
29	265.5	5.5	2109	1 PGCA_CHICK	P07898 gallus gall
30	265	5.5	995	1 YI09_YEAST	P40442 saccharomyc
31	264.5	5.4	881	1 YJH8_YEAST	P47033 saccharomyc
32	263.5	5.4	2700	1 ZAN_HUMAN	Q9y493 homo sapien
33	258.5	5.3	636	1 YNR6_YEAST	P53882 saccharomyc

34 257.5 5.3 860 1 CHI2_COCIM P54197 coccidioide
35 255.5 5.3 676 1 MUC1_MESAU Q00528 mesocricetu
36 255 5.2 605 1 YHC8_YEAST P38739 saccharomyc
37 255 5.2 2437 1 NOTC_BRARE P46530 brachydanio
38 252 5.2 2531 1 NTC1_MOUSE Q01705 mus musculu
39 251 5.2 5376 1 ZAN_MOUSE O88799 mus musculu
40 249.5 5.1 1150 1 APMD_PIG P12021 sus scrofa
41 249 5.1 2531 1 NTC1_RAT Q07008 rattus norv
42 246.5 5.1 634 1 HWPI_CANAL P46593 candida alb
43 246.5 5.1 2703 1 NOTC_DROME P07207 drosophila
44 244 5.0 862 1 PGCV_MACNE Q28858 macaca neme
45 243 5.0 536 1 YEN1_SCHPO O13695 schizosacch

ALIGNMENTS

RESULT 1

AMYH_YEAST
ID AMYH_YEAST STANDARD; PRT; 1367 AA.
AC P08640; P08068;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glucoamylase S1/S2 precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase) (1,4-alpha-D-glucan glucohydrolase).
DE STAL OR STA2 OR MAL5 OR YI019C.
GN Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RA Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RL [2]
RN SEQUENCE OF 1-242 AND 762-1331 FROM N.A.
RX MEDLINE=87194600; PubMed=3106330;
RY Yamashita I., Nakamura M., Fukui S.;
RT "gene fusion is a possible mechanism underlying the evolution of STAL.";
RL J. Bacteriol. 169:2142-2149(1987).
RN [3]
RP SEQUENCE OF 1-31 FROM N.A.
RC STRAIN=SPX101-1C;
RX MEDLINE=89031230; PubMed=3141213;
RA Pardo J.M., Ianez E., Zalacain M., Claros M.G., Jimenez A.;
RT "Similar short elements in the 5' regions of the STA2 and SGA genes from Saccharomyces cerevisiae.";
RL FEBS Lett. 239:179-184(1988).
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-glucose residues successively from non-reducing ends of the chains with release of beta-D-glucose.
CC -1- SIMILARITY: TO S.POMBE SPCC215.13.
CC -1- SIMILARITY: SOME, TO S.POMBE SPCC285.13C.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed, usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

CC -----
CC DRMBL; Z38061; CAA86176.1; -;
CC EMBL; M16164; AAA35014.1; -;
CC EMBL; M16165; AAA35015.1; -;
CC -----

```
DR EMBL; X13857; CAA32069.1; -.
DR PIR; B26877; A26877.
DR PIR; A26877; A26877.
DR PIR; S48478; S48478.
DR SGD; S0001458; MUCL.
KW Hydrolyase; Glycosidase; Polysaccharide degradation; Glycoprotein;
KW Signal; Multigene family.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1367 GLUCOAMYLASE S1/S2.
FT DOMAIN 210 1367 SER/THR-RICH.
FT CARBOHYD 817 817 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 874 874 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1367 AA; 136110 MW; 91C00E2DBD61AA9D CRC64;

Query Match 8.2%; Score 399; DB 1; Length 1367;
Best Local Similarity 24.3%; Pred. No. 2.5e-12;
Matches 196; Conservative 117; Mismatches 325; Indels 170; Gaps 30;

QY 4 TETVRSVAPMRGGEITAHLLTNSTT---SADVT-----GSSASYPEGYNASVLTQFSD 55
DB 400 STTESSAP-----VTSSTESSAPVTSSTTESSAPVTSSTTESSAPVTS 448

QY 56 STVQSGSHALGDRSYSSSS-----TSSSELSNAPRGERSIAGISYQVGRGTAEQR 111
DB 449 STTESSAPVPTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVTSSTESS 508

QY 112 TSSDHTDHTYLSFTFKGERA-----LLSITDNSS---SSDIVESFTSYIKINSHSEY 163
DB 509 SAPVPTP-----SSSTESSAPAPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTES 564

QY 164 SS--FSHAQTERNSISSYDGEYAQPTSPVLHTSNLPSTYPTPIN-MPNTSVVLDTDAEF 220
DB 565 SPTPVTSSSTESSA-----PVPTPSSSTTESSAPVPTPSSSTTESSAPAPTPSSS 617

QY 221 VSDSSSSSSSSSSSSSGPPLPSPVQSQHLLFSSILPSTRASVHLLKSTSDASTPWSSS 280
DB 618 TTSSSAPVPTSTTESSAPVPTPSSS-----TTSSSAPVPTPSSSTTESSA 666

QY 281 PSLPVPVSLTT-SYSAPLSVSQTLLPOSSSTPLPRARET-----PVTSFOTSTMTSPMT 334
DB 667 PVPTPSSSTTESSAPVTSSTT---ESSAPVTSSTTESSAPVPTPSSSTTESSAPVPTP 723

QY 335 LHSQADLKSQSTPHQEKVITESKPSLSVLPTESTKA-----VTTNSPLP--- 381
DB 724 TPSSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVPTP 783

QY 382 -----PSLTESSTQTL-----PATSNLAQMSPTFTTILKTSQLMTPGTIL 425
DB 784 SSTTESSAPVPTPSSSTTESSAPVPTPSSSNITSSAPSPFPSSSTESSVPVPTPS 843

QY 426 SSTASLVGTGPIAVQTTAGKQLSLTHPEILVPQISTEGGISTERNRVIVDATTGLIPLTSV 485
DB 844 SSTTESSAPVSSSTT-----ESSVAPVPTPSSSN-----ITSSAPSIIPSSST 888

QY 486 PTSAKEMTKLGVTAEYSPASRLSGTSPSQTTW----- 520
DB 889 TESFSTGTTVTPSSKYPGSGQTETSVSSTETTIVPTKTTTSVTPSTTTITTTVCSTGT 948

QY 521 ----STAEDLAPKSAFVAVOSSQTPTLLSSASVNSCAVNPCLHNGECVADNTSRGYHC 576
DB 949 NSAGETTSKSPKTVTTVPTTTSSTTSSTTSSTTTTITTVCS-----TGTSAGETTS- 1001

QY 577 RCPSPS--WOGDDCSVDVNECLNCPRCPTATCNNTQGSFICKCPV---GYOLE-KGICNLV 630
DB 1002 -CSPKITTTVPGCSTSPSEFASE---STTSPPTPTVTVVSTTVVTEYSTKPGGEIT 1057

QY 631 RTVFTEFKLRTFLNTVVERHSLDQEVEN--EITKTLNMFSAI-----PSYIR 677
DB 1058 TTEVTK-NIPTVL-TTIAFTPTSVTTVNTFTPTTITTVTCSTGNSAGETTSKCSPKTVT 1115

QY 678 STVHASRESNAVVISLQTTFSLASNVTL 705
DB 705
```

```
DB 1116 TTVPCTGTGTGEYTTTATTLVTTAVTTTV 1143

RESULT 2
ID YN96_YEAST STANDARD; PRT; 1140 AA.
AC Q04893;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 113.1 kDa protein in PRES-FE14 intergenic region.
GN YMR317W OR YM924.09.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Churcher C.M., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z54141; CAA90835.1; -.
DR SGD; S0004936; YMR317W.
KW Hypothetical protein; Repeat.
SQ SEQUENCE 1140 AA; 113070 MW; 0153EBCA24FE5427 CRC64;

Query Match 8.1%; Score 396; DB 1; Length 1140;
Best Local Similarity 21.7%; Pred. No. 2.8e-12;
Matches 230; Conservative 161; Mismatches 387; Indels 284; Gaps 36;

QY 1 MSQTEVRSVAPMRGGEITAHLLTNSTTSADVTG-----SSASYPEG-VNASVLQ 52
DB 52 VSSSTTLVSVWP-----EFTSSSLSDTTASILSSELSVIFSLSYTSSDISSTVND 107

QY 53 FSDSTVQSGSHALGDRSYSSSSSSSLSNLSA-----PRGERSIAGISYQVGRGTA 107
DB 108 VESTSGPSNSYALSSTNAQLSSSTTETDISSAIQTSSPQTSSNNGSGSSEPLGKS 167

QY 108 IEQRTSDHTDHTYL-SSTFKGERALLSITDNSSSDIVESFTSYIKINSHSEYSSF 166
DB 168 SVLETTASSDHTTAVTSSTFT-----TLTDVSSSPKISSSGSAVTSVGTSDASKEVF 220

QY 167 SHAQTERNSISSYDGEYAQPTSPVLHTSNLPSTYPTINMPNTSVVLDTDAEFVSD--- 223
DB 221 SSSTSDVSSLLS-----STSSPA-----SSTSETLPFSTILSTISSVSEAP 265

QY 224 -----SSSSSSSSSSSSSGPPLPSPV-----SQSHLFSIL 257
DB 266 SATSSSVSSSEASSSTSSSVSEAPLATSSVVSVSEAPSTSSVSVSEAPSTSSSVSEIS 325

QY 258 PSTRASV-----HLKSTSDASTPWS-----SSPPLPVSLTTSAP 295
DB 326 STTSSSVSEAPLATSSVVSEAPSTSSSVSSSISSTTSSSVSEAPLATSSVVSEAP 385

QY 296 LSVSQTT---LPGSSSTPV---LPRARETPTVSTQSTMTSFTMLHSSQTADLKSGSTP 349
DB 386 SSTSSSVSEAPSTSSSVSEAPSTSSSVSSSISSTTSSSVSEAPSTSSSVSEAP 445

QY 350 HQEKVITESKPSLSVLPTESTKAVTNTSNPLPPSLTESSTETLPTATSNLAQMSPTFTT 409
DB 446 SAISSLASSRLFSKSKNTSVTSLVATEASSVTSSLRPSS--ETLASNIIESSLSTGYS 503
```


Db 731 LSSNNMSEFTLSTFTVQSIVSFSMEPTSSVASFSSPLLVSRRNSCNDARSSNTISS 790
QY 692 SLOTTFSLASNVF-LFLADRMQKVCNCKSSAEVCLLQSQRIFRAGSLCKRKSPEC 750
Db 791 GLFSTIENVRNATSTNLSTDEIVTSSKSS-----CT 824
QY 751 KDTISICTDLGVALCQCKSGYFQFNKMDHSCRACEDGY----- 788
Db 825 NEDSVLTKT-----QVSVETITSCGGICTLMSPVTTINAKANTLTTE 871
QY 789 -RLENETMCPGGLGNGPNQLITVIAAAGGLLILGIALIVTCCRKNKNDIS 846
Db 872 TSTVETITTCPGGV-----CSTLTVPVTTITSEA-----TTTATISCEDNEEDIT 917

RESULT 4
C114_MOUSE STANDARD; PRT; 573 AA.
AC P19467;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cell surface antigen 114/A10 precursor.
GN 1y64.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89197960; PubMed=2784793;
RA Dougherty G.J., Kay R.J., Humphries R.K.;
RT "Molecular cloning of 114/A10, a cell surface antigen containing
RT highly conserved repeated elements, which is expressed by murine
RT hemopoietic progenitor cells and interleukin-3-dependent cell
RT lines.";
RL J. Biol. Chem. 264:6509-6514(1989).
CC -1- FUNCTION: MAY HAVE A POSITIVE REGULATORY ROLE IN THE CELLULAR
CC RESPONSE TO IL-3.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: THE SER/THR-RICH TANDEM REPEATS MAY SERVE AS SITES OF
CC EXTENSIVE GLYCOSYLATION.
CC -1- PTM: THE MOTIF SER-GLY CONSERVED IN ALL 8 SER/THR-RICH REPEATS MAY
CC SERVE AS THE SITE OF ATTACHMENT OF GLYCOSAMINOGLYCAN SIDE CHAINS.
CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 SEA DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04634; AAA37239.1; -.
DR PIR; A33533; A33533.
DR MGD; MGI:103190; 1y64.
DR InterPro; IPR000561; EGF-like.
DR Pfam; PF01390; SEA; 1.
DR SMART; SM00181; EGF; 1.
DR SMART; SM00001; EGF like; 1.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS50024; SEA; 1.
KW Glycoprotein; Signal; Antigen; EGF-like domain; Repeat; Transmembrane.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 573 CELL SURFACE ANTIGEN 114/A10.
FT DOMAIN 18 480 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 481 508 POTENTIAL.

FT DOMAIN 509 573 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 17 230 8 X TANDEM REPEATS, SER/THR-RICH.
FT REPEAT 17 43 1.
FT REPEAT 44 70 2.
FT REPEAT 71 97 3.
FT REPEAT 98 124 4.
FT REPEAT 125 151 5.
FT REPEAT 152 178 6.
FT REPEAT 179 205 7 (NEAR PERFECT).
FT REPEAT 206 230 8 (APPROXIMATE).
FT DOMAIN 233 273 EGF-LIKE 1.
FT DOMAIN 274 384 SEA.
FT DOMAIN 385 425 EGF-LIKE 2.
FT DISULFID 425 467 EGF-LIKE 3.
FT DISULFID 237 248 BY SIMILARITY.
FT DISULFID 242 257 BY SIMILARITY.
FT DISULFID 259 272 BY SIMILARITY.
FT DISULFID 389 402 BY SIMILARITY.
FT DISULFID 394 408 BY SIMILARITY.
FT DISULFID 410 424 BY SIMILARITY.
FT DISULFID 429 441 BY SIMILARITY.
FT DISULFID 433 451 BY SIMILARITY.
FT DISULFID 453 466 BY SIMILARITY.
FT CARBOHYD 266 266 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 397 397 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 573 AA; 58701 MW; 1154C4F04E2D58A9 CRC64;

Query Match 7.0%; Score 342.5; DB 1; Length 573;
Best Local Similarity 23.8%; Pred. No. 5.1e-10;
Matches 161; Conservative 73; Mismatches 230; Indels 213; Gaps 26;
QY 221 VDSOSSSSSSSSSSSSSSGPPPLPSPVSQSHHLPFSSILPSTRASVHLKSTSDASTPWSSS 280
Db 16 VGSSSQASSTSSSGGSPPTTVQSQSPG-----SSQASSTTSSS 56
QY 281 PSLPVLSTTSAPLSVSTQTLTLPSSSTPVLPRARETPTVTSFQTSTMTSPTMLHSSQT 340
Db 57 GGASPTT-TVQSQSPGSSSQASTTSSGGASP-----PTT----- 91
QY 341 ADLKSQSTPHQKIVITESKPSLSVSLPTESTKAVTTHSPLPPLTESSTEQTLPATSNL 400
Db 92 --VQSQSPGSSSQASTTSSSGGASPPT-----TVQSQSPGSSSQAST-----TTSSS 137
QY 401 AQMSPTTTTILKTSQPLMTTPGTLSTASLVTGPIAVQTAGKQLSLTHPEILVPOIST 460
Db 138 GGASP--PTTVQSQS-----PGS--SSQASTTT-----SS 163
QY 461 EGGISTERNRIVDATTGLIPLTSVPTSAKEMTTKLGVTAEYSPASRSLGTSPTQPTTV 520
Db 164 SGGAS-----PPTTVQSQSP-----GSSSQVSTTSSSG-GASPTTVQ 201
QY 521 STADLAPKATFATVQSQSTOSPTLLSSASVNSCAVNPCLHNGECVADNTRSGVHCRCPP 580
Db 202 SQ-----SPGS-----SSQPGTQPSGGASST-----VPSGGSTGSDLCNP 239
QY 581 SWGDDCSVDVNECLSNPCPSTATCNTQSGFICKCPVGOLEKICNLVTRFTVEFKLK 640
Db 240 -----NPKGTASCVKLHSHKHFCLCEGYIYNSLSSCVK-----G 275
QY 641 RTFLNTTVERKHSIDLQEVENEITKTLNMCFSALPSYIRSTVHASRESNAVVISLQTTFSLA 700
Db 276 TTPGDISMSVSETANLEDENSVGQELYSVTDFFETFNKTDYQGTVILKVSATPSRS 335
QY 701 SNVTFLFLADRMQKV-----NSCKSSAEVCLLQSQRRIFRAGSLCKRKSPE 748
Db 336 ARSAMRATKDVSVSVYVNIFFGADTKETEKSVSSAIAETKTSGNVKDYVSNLCHDYG-- 393
QY 749 C-DKDTISICTDLGVALCQCKSGYFQFN-----KMDHSCR- 782
Db 394 CVGNDSKQCD----ILOCTCKPLGLRLNPQVPCFVAVTCQPCNAEKEQCLKMDNGVMD 450


```

QY 783 -ACEDGYRLE--NETMSCPFGGLGNCNPNYQLITVIVAAAGGGLLLILGIALIVTCR 839
Db 451 CVCMPGYORANGNRKCECPGYSGMCKDQFQLILVITGVIAGALLILLIAFVSARS 510
QY 840 KNKNDISKLFKSGDFQ 856
Db 511 KNKK-----KDGEQ 520

RESULT 5
PGCN_RAT
ID PGCN_RAT STANDARD; PRT: 1257 AA.
AC P55067;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-OCT-2001 (Rel. 40, Last annotation update)
DE Neurocan core protein precursor (245 kDa early postnatal core
DE glycoprotein) [Contains: 150 kDa adult core glycoprotein].
GN CSPG3 OR NCAN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RA MEDLINE=92406907; Pubmed=1326557;
RT "Cloning U., Karthikeyan L., Maurel P., Margolis R.U., Margolis R.K.;
RT neuronal cell adhesion molecule of neurocan, a developmentally
RT regulated, aggregating chondroitin sulfate proteoglycan of brain.";
RL J. Biol. Chem. 267:19536-19547(1992).
RN [2]
RN CHARACTERIZATION.
RX MEDLINE=94230574; Pubmed=7513709;
RA Friedlander D.R., Milev P., Karthikeyan L., Margolis R.K.,
RA Margolis R.U., Grumet M.;
RT "The neuronal chondroitin sulfate proteoglycan neurocan binds to the
RT neural cell adhesion molecules Ng-CAM/L1/NILE and N-CAM, and inhibits
RT neuronal adhesion and neurite outgrowth.";
RL J. Cell Biol. 125:669-680(1994).
CC -1- FUNCTION: MAY MODULATE NEURONAL ADHESION AND NEURITE GROWTH DURING
CC DEVELOPMENT BY BINDING TO NEURAL CELL ADHESION MOLECULES (NG-CAM
CC AND N-CAM). CHONDROITIN SULFATE PROTEOGLYCAN; BINDS TO HYALURONIC
CC ACID.
CC -1- TISSUE SPECIFICITY: EARLY POSTNATAL AND ADULT BRAIN; NOT EXPRESSED
CC IN KIDNEY, LUNG, LIVER AND MUSCLE.
CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO N-LINKED AND
CC O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
CC -1- PTM: TWO ISOFORMS WERE FOUND THAT PROBABLY ARISE BY PROTEOLYTIC
CC DEGRADATION. THE LARGE ISOFORM IS PREDOMINANT IN EARLY POSTNATAL
CC BRAIN, THE SMALL ISOFORM IS FOUND IN ADULT BRAIN.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M97161; AAC37679.1; -.
CC HSP: P00740; IIXA.
CC InterPro: IPR000152; Asx_hydroxyl.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR000742; EGF-2.
CC InterPro: IPR001881; EGF_Ca.
CC InterPro: IPR003599; Ig.

```

```

DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR000538; Link.
DR InterPro: IPR000436; Sushi_SCR_CCP.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00008; EGF; 2.
DR Pfam: PF00047; Ig; 1.
DR Pfam: PF00059; lectin_c; 1.
DR Pfam: PF00084; sushi; 1.
DR Pfam: PF00193; Xlink; 2.
DR ProDom: PD000918; Link; 2.
DR SMART: SM00032; CCP; 1.
DR SMART: SM00034; CLECT; 1.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00001; EGF_like; 1.
DR SMART: SM00409; IG; 1.
DR SMART: SM00445; LINK; 2.
DR PROSITE: PS00010; ASX_HYDROXYL; 1.
DR PROSITE: PS00022; EGF_1; 3.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS01241; LINK; 2.
DR PROSITE: PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE: PS50041; C_TYPE_LECTIN_2; 1.
KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1257 NEUROCAN CORE PROTEIN.
FT CHAIN 639 1257 150 KDA ADULT CORE GLYCOPROTEIN.
FT DOMAIN 51 146 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 176 253 LINK 1.
FT DOMAIN 274 355 LINK 2.
FT DOMAIN 949 985 EGF-LIKE 1.
FT DOMAIN 987 1023 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 1025 1154 C-TYPE LECTIN.
FT DOMAIN 1155 1213 SUSHI.
FT DISULFID 58 139 BY SIMILARITY.
FT DISULFID 181 252 BY SIMILARITY.
FT DISULFID 205 226 BY SIMILARITY.
FT DISULFID 279 354 BY SIMILARITY.
FT DISULFID 303 324 BY SIMILARITY.
FT DISULFID 953 964 BY SIMILARITY.
FT DISULFID 958 973 BY SIMILARITY.
FT DISULFID 975 984 BY SIMILARITY.
FT DISULFID 1029 1040 BY SIMILARITY.
FT DISULFID 1057 1149 BY SIMILARITY.
FT DISULFID 1125 1141 BY SIMILARITY.
FT DISULFID 1156 1199 BY SIMILARITY.
FT DISULFID 1185 1212 BY SIMILARITY.
FT CARBOHYD 121 121 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 339 339 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 737 737 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 944 944 O-LINKED (XYL. . .) (CHONDROITIN SULFATE).
FT CARBOHYD 967 967 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1164 1164 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1257 AA; 135544 MW; 992B33DCFA19EE1B CRC64;

```

Query Match 6.88; Score 328; DB 1; Length 1257;

Best Local Similarity 22.08; Pred. No. 6.2e-09; Matches 155; Conservative 107; Mismatches 255; Indels 186; Gaps 31;

QY 1 MSQTETYSRSVAPMRGGEITAHLLTNTSTTSADVTGSSA-----SYPEGVNASVLTFQ 53

Db 418 LTRQASQETLSTPPGGTTLASWLLTGTSTSGVPSPSSLGVDMEETPSG-----TQV 471

QY 54 SDSTVQSGGSHALGDRSYSESSSTSSSESLNSAPRGERSIAGISYGVQVRGTAEIORTS 113

Db 472 APTPTMRGRFKGLNGRHFQOQG--PEDQLLEAAEAQAQPTPLEVTADHMGPSAATALE 529

QY 114 SDHTDHTYLSSTFTTKGERALLSITDNSSSDIVESSYIKISNSSHSEYSSFSHAQTER 173

Db 530 SDOSHSPW-----AIL-----TNEVDVPGAGS-----LGRSLPESRKWSPSLISP 570

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Brain;
RX MEDLINE=96039250; PubMed=7490074;
RA Rauch U., Grimpe B., Kuibe G., Arnold-Ammer I., Beier D.,
RA Fessler R.;
RA "Structure and chromosomal localization of the mouse neurocan gene.";
RL Genomics 28:405-410(1995).
CC -!- FUNCTION: MAY MODULATE NEURONAL ADHESION AND NEURITE GROWTH DURING
CC DEVELOPMENT BY BINDING TO NEURAL CELL ADHESION MOLECULES (NG-CAM
CC AND N-CAM). CHONDROITIN SULFATE PROTEOGLYCAN; BINDS TO HYALURONIC
CC ACID.
CC -!- TISSUE SPECIFICITY: BRAIN.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE AGGRECAM/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X84727; CAA59216.1; .
DR HSP; P00740; LEDM.
DR MGD; MGI:104694; Cspg3.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR001304; lectin_c.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00193; Xlink; 2.
DR ProDom; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00179; EGF_CA; 1.
DR SMART; SM00001; EGF_like; 1.
DR SMART; SM00409; IG; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00022; EGF_1; 3.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
DR PROSITE; PS00615; C-TYPE-LECTIN.1; 1.
DR PROSITE; PS50041; C-TYPE-LECTIN.2; 1.
KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW EGF-like domain; Calcium; Repeat; Lectin; Sushi; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1268
FT DOMAIN 51 146
FT DOMAIN 176 253
FT DOMAIN 274 355
FT DOMAIN 960 1034
FT DOMAIN 998 1036
FT DOMAIN 1036 1165

FT	DOMAIN	1166	1224	SUSHI.
FT	DISULFID	58	139	BY SIMILARITY.
FT	DISULFID	181	252	BY SIMILARITY.
FT	DISULFID	205	226	BY SIMILARITY.
FT	DISULFID	279	354	BY SIMILARITY.
FT	DISULFID	303	324	BY SIMILARITY.
FT	DISULFID	964	975	BY SIMILARITY.
FT	DISULFID	969	984	BY SIMILARITY.
FT	DISULFID	1040	1051	BY SIMILARITY.
FT	DISULFID	1068	1160	BY SIMILARITY.
FT	DISULFID	1136	1152	BY SIMILARITY.
FT	DISULFID	1167	1210	BY SIMILARITY.
FT	DISULFID	1196	1223	BY SIMILARITY.
FT	CARBOHYD	121	121	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	339	339	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	742	742	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	978	978	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1175	1175	N-LINKED (GLCNAC. .) (POTENTIAL).
SQ	SEQUENCE	1268 AA; 137200 MW; 3014E8E202A2FAEC CRC64;		

Query Match 6.5%; Score 315.5; DB 1; Length 1268;
Best Local Similarity 23.0%; Pred. No. 2.5e-08;
Matches 164; Conservative 95; Mismatches 259; Indels 195; Gaps 31;

Qy	1	MSQETVSRVAPMRGGEITA-----HWLLTNSTTSADVTGSS-----ASYPEGVNA	47
Db	418	LTWTQAPETLPGTGGPTLASWFSSEKWLFTGAPSMGVSPSDMGVDMETPLG---	474
Qy	48	SVLQFSDTSVQSGSHPALGDRSVSESSSTSSSLNASSAPRGERSIAGISYGVGRGTA	107
Db	475	---TQVAPTPTMRGRFRGLNGRHFOQGPEDQLPEV--AEPQAOPPLGATANIMRPSA	529
Qy	108	IEQRTSSDHTYLSSTFTGERALLSITDNSSSDIVSESTSYKISNSHSHSYSFS	167
Db	530	ATEASESDQSHSPWAILTNEVDPCAGSLGSRSLPESLMWSPS---LISPSVPS-----	580
Qy	168	HAQTERSNISSYDGEYAQPSTESPVLHTSNLPSTPTTNMNTSVVLTDAEFVSDSSSS	227
Db	581	---TE-STPSPKPGAAEAPSVKSAIPLRLPSEPPA---PSPG-----PSEALSA	624
Qy	228	SSSSSSSSSGPPLPLPSVSQSHLFF---SSILPSTRASVHLKSTSDASTPWSSP-S	282
Db	625	VLQASSADGSPDPPIVAMLRAPKLWLLPRSTLVPN-----LHSSQTA-----	670
Qy	283	PL-----PVSL-----TTTSAPLSVSQTTLPQSSSTPV---LPRARET	318
Db	671	PLPSWVPEQAVRPVSLGAEDLETPFTTIAAPVEASHRS-PDADSIEIGTSSMRATKH	729
Qy	319	PVTSEQTSTWTSFMTM-----LHSSQTA-----	341
Db	730	PISGFWASLDSNVTMPVPSPDAGILGTESGVLDLPGSTGGQATVEKVLATWLPQCG	789
Qy	342	--DLKSQSTPHQKIVITESKPSLVSL-----PTESTKAVTNTSPLPPSLSTESSTQ	392
Db	790	GLDPSQSTPMEANGVAVSMEPT-VALEGGATEGMEATREVU-----PSTADATWE--	840
Qy	393	LPATSTNLAQMSPTFTTTILKTSQPLMTTPGTLSTASLVTGPIAVOTTAGKQLSL-THP	451
Db	841	---SESRSAISSTHIAVTMARACGMPTLTSSBSGHPGPKGMVQAQESLEPLNTLPSP	896
Qy	452	--EILVPOISTEGGISTERNRIVDATTGL--IPLTSVPTSAKEMTTKLGVTAEVSPASR	507
Db	897	WSSLVVPMDVEVASVSGE-----PTGLWDIPSTLIPVSLGLDESVLNVAA-----	942
Qy	508	SLGTSPSPQTTVSTAEADLPKSAFTAVQSSQSTQSTPTTLLSSASVNSCAVNCPLHNGE	567
Db	943	----SPS-----VEGFWEVVA-----SGQEDPT-----DPCENNPLHGTC	977
Qy	568	DNTSRGYHCRPPSPWQGDGCSVDVNECLSNPCPSTATCNNTQGSFICKPCPVGY	620
Db	978	NGTV--YCCSCDQYAGENCEIDDDCLCSPCENGCTCIDVNGFICLCILFSY	1028

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>) or send an email to license@sib-sib.ch.

CC EMBL; X60226; CAA42787.1; -
 CC EMBL; D13542; BAA02742.1; -
 CC HSSP; P20693; 1HLJ
 CC InterPro; IPR00152; Asx_hydroxyl.
 CC InterPro; IPR00561; EGF-like.
 CC InterPro; IPR00742; EGF_2.
 CC InterPro; IPR01881; EGF_Ca.
 CC InterPro; IPR003599; Ig.
 CC InterPro; IPR003006; Ig_MHC.
 CC InterPro; IPR00538; Link.
 CC InterPro; IPR00436; Sush1_SCR_CCP.
 CC InterPro; IPR001304; lectin_C.
 CC Pfam; PF00008; EGF; 2.
 CC Pfam; PF00047; Ig; 1.
 CC Pfam; PF00059; lectin_c; 1.
 CC Pfam; PF00084; sushi; 1.
 CC Pfam; PF00193; Xlink; 2.
 CC ProDom; PD000918; Link; 2.
 CC SMART; SM00032; CCP; 1.
 CC SMART; SM00034; CLECT; 1.
 CC SMART; SM00179; EGF_Ca; 1.
 CC SMART; SM00001; EGF_Like; 1.
 CC SMART; SM00409; IG; 1.
 CC SMART; SM00445; Link; 2.
 CC PROSITE; PS00010; ASX_HYDROXYL; 1.
 CC PROSITE; PS00022; EGF_1; 2.
 CC PROSITE; PS01186; EGF_2; 1.
 CC PROSITE; PS01187; EGF_Ca; 2.
 CC PROSITE; PS01241; LINK; 2.
 CC PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
 CC PROSITE; PS00041; C_TYPE_LLECTIN_2; 1.
 CC Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
 KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
 KW Hyaluronic acid; Alternative splicing.
 FT SIGNAL 1 26 POTENTIAL.
 FT CHAIN 27 3562 VERSICAN CORE PROTEIN.
 FT DOMAIN 37 136 IG-LIKE V-TYPE DOMAIN.
 FT DOMAIN 166 243 LINK 1.
 FT DOMAIN 264 345 LINK 2.
 FT DOMAIN 3254 3290 EGF-LIKE 1.
 FT DOMAIN 3292 3328 EGF-LIKE 2. CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 3341 3455 C-TYPE LECTIN.
 FT DOMAIN 3460 3518 SUSHI.
 FT DISULFID 44 129 BY SIMILARITY.
 FT DISULFID 171 242 BY SIMILARITY.
 FT DISULFID 195 216 BY SIMILARITY.
 FT DISULFID 269 344 BY SIMILARITY.
 FT DISULFID 293 314 BY SIMILARITY.
 FT DISULFID 3258 3269 BY SIMILARITY.
 FT DISULFID 3263 3278 BY SIMILARITY.
 FT DISULFID 3280 3289 BY SIMILARITY.
 FT DISULFID 3296 3307 BY SIMILARITY.
 FT DISULFID 3301 3316 BY SIMILARITY.
 FT DISULFID 3318 3327 BY SIMILARITY.
 FT DISULFID 3334 3345 BY SIMILARITY.
 FT DISULFID 3362 3454 BY SIMILARITY.
 FT DISULFID 3430 3446 BY SIMILARITY.
 FT DISULFID 3461 3504 BY SIMILARITY.
 FT DISULFID 3490 3517 BY SIMILARITY.
 FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 235 235 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 329 329 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 709 709 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 948 948 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 1409 1409 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1479 1479 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1523 1523 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1530 1530 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1625 1625 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1751 1751 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1988 1988 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2088 2088 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2089 2089 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2507 2507 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2642 2642 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2679 2679 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2748 2748 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 2762 2762 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3069 3069 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3194 3194 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3232 3232 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 3545 3545 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 485 1411 MISSING (IN ISOFORM V1).
 SQ SEQUENCE 3562 AA; 388078 MW; 9BC566E88C1602D2 CRC64;
 Query Match 6.48; Score 309.5; DB 1; Length 3562;
 Best Local Similarity 22.4%; Pred. No. 1.6e-07;
 Matches 160; Conservative 93; Mismatches 255; Indels 205; Gaps 30;
 QY 26 TNSTADVTGSSASYPEGV-----NASV-LTQFSDSTVQS-GGSH TALGDRSYSESS 76
 DB 2697 SEKTIIDM-DHKSMPEDIIISVOTMPLVIRSTQVSDNNKAEEDKYDLSINFSTEEN 2755
 QY 77 STSSSESLN---SSAPRGERSIAG-----ISYGVGRGTAIEQRTS----- 113
 DB 2756 SFGSGDNLSTTSIOPSSSVTAGCHKVLDKGLGSGYAMQFATETLTITVLNGLFLP 2815
 QY 114 -----SDHTDHTYLSSTFTKGERALLSITDNSSSDIVESSTSYIKISNSHSEYSSFS 167
 DB 2816 TVPSLVSPHMPHESKESEFEAKHIGRTSTD-----DYIEPYTS---ANNOVITD----- 2862
 QY 168 HAQTERENISYDGEYAPQSTESVPLHTSNLPSYTPINNMPTSIVLTDIAEFYSDSSSS 227
 DB 2863 --QSKTWSISGFGSGMOEGSGDKKPM---IPSLTPDLTM-ETEKALTTDFDFVSWYTO 2915
 QY 228 SSSSSSSSGGPPPLPSVSQSHLFFSILPSTRASVHLKSTSDAS-----T 275
 DB 2916 SMSQATVSS-----SSSEKHTSVYMQTKASATEYEEDTSVLSNSVQNPCKSSVT 2966
 QY 276 PW-SSSPSPLPVSLTSTTSAPLSVSOT-----TLPOSSSTPPVLPARETPTVSFQISTM 328
 DB 2967 VWLVNGVSKYPEVIIPSTSSAKDSQSDHSDSGTTFKEVSSD--MAATYKPTTDLDT-TV 3023
 QY 329 TSFMTLHSSQTADLKQSOTPHQEKVIT---ESKSPSLVSLPTSTKAVTNSPLPPLSLT 385
 DB 3024 SLLLVFSPPESESISTESTPHFNKFVTERSEETESSVNDLIIENATVSGDSP----- 3077
 QY 386 ESSTEQTLPATSNLAQMSFTFTTILKTSOPLMTTGTLSSTASLTGPIAVOTTAGKQ 445
 DB 3078 ---SIHDYPTAFWNFGERT-----STDVPKLST-----IEVEFSSER- 3111
 QY 446 LSLTHPEILVPQISTEGGISTERNRVIVDATGLIPLTSVPTSAAKMTTKLV----- 498
 DB 3112 -----VKNPQESDRSTEREP-----RUSSAPVSDSPNSIEVGVFKPQDEA 3153
 QY 499 ---TAEYSPASRLGTSPS-----PQTTVVYST---AEDLAPKATFAVOSSTOSPT 543
 DB 3154 VTMLTSSLEPLDRSLETQSALLGPLLQGOEITITISSNIATNNTAPGNPNYNEQSTISSE 3213
 QY 544 TSSSSASVNS-----CAVNPCLHNGECVADN 569
 DB 3214 LLNTIELVTSFSLPEVTNGSDFLIGTSVSGVEGTAVOIQDPCKSNPCLNGTCTY--- 3270
 QY 570 TSGR--YHCRCPSPWGDDCSVDVNECLSNPCPSTACNNQTGCSFKICKCPGV 620
 DB 3271 -PRGSFYICLPLFGNGEQCELDIDECQSNPCRNAGATCIDGLNTFTTCLCLPSY 3322

```

RESULT 11
ID AGAL_YEAST STANDARD; PRT; 725 AA.
AC P32323;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE A-agglutinin attachment subunit precursor.
GN AGAL OR YNR044W OR N3431.
GS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=91304412; Pubmed=2072914;
RA Roy A., Lu C.F., Marykwas D.L., Lipke P.N., Kurjan J.;
RT "The AGAL product is involved in cell surface attachment of the
RT Saccharomyces cerevisiae cell adhesion glycoprotein a-agglutinin.";
RN Mol. Cell. Biol. 11:4196-4206(1991).
[2]
RN SEQUENCE FROM N.A.
RP Pohl T.M.;
RA Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MEDIATES THE CELL SURFACE ATTACHMENT OF THE A-AGGLUTININ
CC SUBUNIT. S.CEREVISIAE A AND ALPHA CELLS EXPRESS THE COMPLEMENTARY
CC CELL SURFACE GLYCOPROTEINS A-AGGLUTININ AND ALPHA-AGGLUTININ,
CC RESPECTIVELY, WHICH INTERACT WITH ONE ANOTHER TO PROMOTE CELLULAR
CC AGGREGATION DURING MATING.
CC -!- SUBUNIT: CONTAINS AT LEAST A BINDING SUBUNIT DISULFIDE-LINKED TO
CC A CORE SUBUNIT.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC AND/OR BY CARBOHYDRATE-MEDIATED COVALENT CROSS-LINKS (POSSIBLE).
CC -!- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M60590; AAA34382.1; -
CC EMBL; Z71659; CAA96325.1; -
CC PIR; S17031; S17031.
CC PIR; A41258; A41258.
CC DR; S0005327; AGAL.
CC DR; S0005327; AGAL.
CC KW Glycoprotein; Cell adhesion; Signal; GPI-anchor; Repeat;
CC Phorbolone response.
CC SIGNAL 1 22
CC FT CHAIN 23 725
CC FT DOMAIN 53 493
CC FT REPEAT 53 149
CC FT REPEAT 395 493
CC FT DOMAIN 182 307
CC FT POTENTIAL.
CC FT A-AGGLUTININ ATTACHMENT SUBUNIT.
CC FT 2 X APPROXIMATE REPEATS.
CC FT 1-1.
CC FT 1-2.
CC FT 18 X APPROXIMATE TANDEM REPEATS, SER/THR-
CC FT RICH.
CC FT REPEAT 182 188
CC FT REPEAT 189 195
CC FT REPEAT 202 202
CC FT REPEAT 203 209
CC FT REPEAT 210 216
CC FT REPEAT 217 223
CC FT REPEAT 224 230
CC FT REPEAT 231 237
CC FT REPEAT 238 244
CC FT REPEAT 245 251
CC FT REPEAT 252 258
CC FT REPEAT 259 265
CC FT REPEAT 266 272
CC FT REPEAT 273 279

```


DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
DE hyaluronate-binding protein) (GAP).
GN CP5G2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS V0; V1; V2 AND V3).
RC TISSUE-Forebrain;
RX MEDLINE-98288320; PubMed-9624174;
RA Schmalfeldt M., Dours-Zimmermann M.T., Winterhalter K.H.,
RA Zimmermann D.R.;
RT "Versican V2 is a major extracellular matrix component of the mature
RT bovine brain.";
RL J. Biol. Chem. 273:15758-15764(1998).
RN [2]
RP SEQUENCE OF 21-53; 78-96; 226-250; 262-277; 295-306; 314-324; 329-331
RP AND 342-348.
RC TISSUE-Spinal cord;
RX MEDLINE-92062692; PubMed-1720020;
RA Perides G., Biviano F., Bignami A.;
RT "Interaction of a brain extracellular matrix protein with hyaluronic
RT acid.";
RL Blochm. Biophys. Acta 1075:248-258(1991).
CC -1- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -1- ALTERNATIVE PRODUCTS: At least 4 isoforms; V0 (shown here), V1, V2
CC and V3; are produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 are expressed
CC in the central nervous system, and in a number of mesenchymal and
CC epithelial tissues; the major isoform V2 is restricted to the
CC central nervous system.
CC -1- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC (by similarity).
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF060456; AAC24358.1; -;
DR EMBL; AF060457; AAC24359.1; -;
DR EMBL; AF060458; AAC24360.1; -;
DR EMBL; AF060459; AAC24361.1; -;
DR HSSP; P01132; 1EGF.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR011881; EGF_Ca.
DR InterPro; IPR003599; Ig.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003600; Ig_like.
DR InterPro; IPR000538; Link.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR001304; Lectin_c.
DR Pfam; PF00008; EGF; 2.
DR Pfam; PF00047; Ig; 1.
DR Pfam; PF00059; lectin_c; 1.
DR Pfam; PF00084; sushi; 1.

DR Pfam; PF00193; xlink; 2.
DR PRINTS; PR01265; LINKMODULE.
DR ProDom; PD000918; Link; 2.
DR SMART; SM00032; CCP; 1.
DR SMART; SM00034; CLECT; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00179; EGF_CA; 2.
DR SMART; SM00409; IG; 1.
DR SMART; SM00410; IG_like; 1.
DR SMART; SM00445; LINK; 2.
DR PROSITE; PS00010; ASX_HYDROXYL; 1.
DR PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR PROSITE; PS00041; C_TYPE_LECTIN_2; 1.
DR PROSITE; PS00022; EGF_1; 2.
DR PROSITE; PS01186; EGF_2; 1.
DR PROSITE; PS01187; EGF_CA; 1.
DR PROSITE; PS01241; LINK; 2.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 3381 VERSICAN CORE PROTEIN.
FT DOMAIN 37 138 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 168 245 LINK 1.
FT DOMAIN 266 347 LINK 2.
FT DOMAIN 349 1336 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT
FT DOMAIN) DOMAIN).
FT DOMAIN 1337 3074 GAG-BETA.
FT DOMAIN 3074 3110 EGF-LIKE 1.
FT DOMAIN 3112 3148 EGF-LIKE 2, CALCIUM-BINDING.
FT DOMAIN 3161 3275 C-TYPE LECTIN.
FT DOMAIN 3280 3338 SUSHI.
FT DISULFID 44 131 BY SIMILARITY.
FT DISULFID 173 244 BY SIMILARITY.
FT DISULFID 197 218 BY SIMILARITY.
FT DISULFID 271 346 BY SIMILARITY.
FT DISULFID 295 316 BY SIMILARITY.
FT DISULFID 3078 3089 BY SIMILARITY.
FT DISULFID 3083 3098 BY SIMILARITY.
FT DISULFID 3100 3109 BY SIMILARITY.
FT DISULFID 3116 3127 BY SIMILARITY.
FT DISULFID 3121 3136 BY SIMILARITY.
FT DISULFID 3138 3147 BY SIMILARITY.
FT DISULFID 3154 3165 BY SIMILARITY.
FT DISULFID 3182 3274 BY SIMILARITY.
FT DISULFID 3250 3266 BY SIMILARITY.
FT DISULFID 3281 3324 BY SIMILARITY.
FT DISULFID 3310 3337 BY SIMILARITY.
FT CARBOHYD 57 57 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 331 331 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 817 817 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 965 965 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1017 1017 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1333 1333 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1393 1393 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1437 1437 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1463 1463 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1653 1653 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 1974 1974 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 2045 2045 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 2074 2074 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 2103 2103 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 2263 2263 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 2290 2290 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 2356 2356 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 2623 2623 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 2641 2641 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 2919 2919 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 3052 3052 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 3354 3354 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 3364 3364 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARSPLIC 349 349 P -> R (IN ISOFORM V1 AND V3).

FT	VARSPUBLIC	350	1336	MISSING (IN ISOFORM V1).
FT	VARSPUBLIC	1337	3074	MISSING (IN ISOFORM V1).
FT	VARSPUBLIC	350	3074	MISSING (IN ISOFORM V3).
FT	CONFLICT	25	25	MISSING (IN REF. 2).
FT	CONFLICT	51	51	MISSING (IN REF. 2).
FT	CONFLICT	89	89	N -> D (IN REF. 2).
FT	CONFLICT	96	96	Q -> D (IN REF. 2).
FT	CONFLICT	346	346	C -> R (IN REF. 2).
SO	SEQUENCE	3381	AA; 369984 MW; F09716FA7778D459	CRC64;
Query Match 6.0%; Score 293; DB 1; Length 3381;				
Best Local Similarity 22.0%; Pred. No. 9.4e-07;				
Matches 149; Conservative 108; Mismatches 274; Indels 146; Gaps 27;				
Qy	1	MSQETVSRSVAPMRGGHITAHWLLTNSTTSDVTGSSASYPE--GVNASVLTQFSDSTV	58	
Db	2555	LTMTSTILEIP---ELT-----SKNTVIDIDHTKPIYEDILGMQTDLDPEVPSGPP	2605	
Qy	59	QSGSHTALGDRSYSESSSTSSSESLNAPPGRGSIAGISYGOVRCGTAEIORTSSDHTD	118	
Db	2606	DSSSESTQVQEKYCAANVLSSTEENFEAS---GDILLANYQATPESKAPEDRNPLDHTD	2662	
Qy	119	HTYLSSTFTGERALLSTTNDSSSDIVESTSYIKI-SNSSHSEYSSFSHAQT-----E	172	
Db	2663	F-----IFTTG-----IPLSSETELDVLLPTATSLPIPSKSATVNPESKTEAKTLEDIFE	2713	
Qy	173	RSNISSVDGEYAQSTSPVLHTNLPSTYPTINMPNTSVVLDTAEPVSDSSSSSSSSSS	232	
Db	2714	SSTLS--DGOAI--ADQSEVISTLGYLERTQN-----EDEAKKYVSPSPQPEFSSG	2760	
Qy	233	SSSSSGPPLPLPSVSOHHLFSSILPSTRASVHLKSTSDASTPKSSSPSPPLPVSLTST	292	
Db	2761	AEALTDPTPVVSGTYI-----LTAQSLTEADPMVEGAPLPSIDIST	2804	
Qy	293	SA-----PLVSQSTTLQPSSSTPVLPRARETPTVTFQSTMTSEMTMLHSQ	339	
Db	2805	VSAPSELLSQTPFPPLSIHLGSGDSEHSDLPSPA--LPSTDASTPPVS-----SGE	2855	
Qy	340	TADLKSOSTPHQEVITESKPSLVSLPTSTKAVTINSPLPPLSTESSTQTLPATSTN	399	
Db	2856	LANIEATFKPSEEDFYITEPP---SLPPDPESEDESKPKLLEPTEASATELIAQEEIE	2912	
Qy	400	LAQMSPTFTTTILKTSQ-----PLMTTP-----GTL---SSTASLVTPG	435	
Db	2913	IFQNSDN-TTSVOVSGEAVKVPFPIETPEAEAIYTAASETKLEGATLRPHSTASVINGV	2971	
Qy	436	IAYOTTAGKQLSLTHPEILVP-QISTEGGISTERNRVIVDATTGLIPLTSVPT-----	487	
Db	2972	EAGVVPQPSQTSRPTLSPLSISPE-----TQAALINGEDSTVAAPKQVPTRMLDN	3027	
Qy	488	---SAKEMTTKLGVTAEYSPASRLGTSPPSPQTVVYSTAEDLAPKSATFAVQSTQSPT	543	
Db	3028	QATLSTTELNTELA-----TPSFPLLETS-NETSFLIGINEESVEGTAVYL-----	3072	
Qy	544	TLSSSASVNSCAVNPCLHNGECVADNTSRGYHCRCPSPSGOQDGCSDVDVNECLSNPCPSTA	603	
Db	3073	-----PGPDRCKMNPCLNGGTCYETEIS--YVCICVGYSGDRCELDEDECHSNPCRNCA	3125	
Qy	604	TCNNTQGSFICKCPVGY	620	
Db	3126	TCIDGNTFRCLCLPSY	3142	

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 14:28:36 ; Search time 41.29 Seconds
(without alignments)

3963.507 Million cell updates/sec

Title: US-09-840-746-1

Perfect score: 4859

Sequence: 1 MSQETVSRVAPMRGGEIT.....FPGQYNPFSIDSERRDYF 946

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3409	70.2	654	Q9ULI3	Q9ul13 homo sapien
2	1315	27.1	245	Q9BSV0	Q9bsv0 homo sapien
3	392.5	8.1	3507	Q23587	Q23587 caenorhabdi
4	387.5	8.0	2232	P91365	P91365 caenorhabdi
5	364	7.5	2271	Q990Y4	Q99gy4 staphylococ
6	363.5	7.5	957	Q14651	Q14651 homo sapien
7	363.5	7.5	1217	Q9URW9	Q9ukw9 homo sapien
8	352.5	7.3	1236	Q9C105	Q9c105 schizosacch
9	348	7.2	1079	Q9N4S7	Q9n4s7 caenorhabdi
10	347	7.1	528	Q29071	Q29071 sus scrofa
11	343.5	7.1	547	P97881	P97881 rattus norv
12	339	7.0	786	Q21027	Q21027 caenorhabdi
13	337.5	6.9	901	Q9H195	Q9h195 homo sapien
14	336	6.9	3072	Q939N5	Q939n5 streptococc
15	334.5	6.9	2178	Q9KWR3	Q9kwr3 streptococc
16	334	6.9	3381	Q9KX33	Q9kx33 streptococc

17	331.5	6.8	3971	3	Q96WV6	Q96wv6 schizosacch
18	331	6.8	1203	5	Q9NSK0	Q9nsk0 caenorhabdi
19	326	6.7	957	4	Q9URN0	Q9urn0 homo sapien
20	325	6.7	849	3	P87107	P87107 saccharomyc
21	323	6.6	4776	16	Q97P71	Q97p71 streptococc
22	321	6.6	878	4	Q9GZ22	Q9gz22 homo sapien
23	319	6.6	878	4	Q9H307	Q9h3q7 homo sapien
24	317.5	6.5	877	4	Q9H306	Q9h3q6 homo sapien
25	317	6.5	1443	3	Q9Y8F2	Q9y8f2 candida alb
26	317	6.5	1795	5	O76894	O76894 drosophila
27	316.5	6.5	1325	5	Q9BKV7	Q9bkv7 leishmania
28	314	6.5	816	11	O70474	O70474 rattus norv
29	313.5	6.5	534	3	Q94317	Q94317 schizosacch
30	313	6.4	1802	3	Q04051	Q04051 saccharomyc
31	310.5	6.4	948	3	O74346	O74346 schizosacch
32	310.5	6.4	3570	4	Q99552	Q99552 homo sapien
33	310	6.4	961	3	Q92223	Q92223 emericella
34	307.5	6.3	2586	5	Q9VTK8	Q9vtk8 drosophila
35	303.5	6.2	1805	11	O63661	O63661 rattus norv
36	301	6.2	3178	5	Q969D4	Q969d4 caenorhabdi
37	300.5	6.2	622	4	Q14881	Q14881 homo sapien
38	300.5	6.2	1275	5	O76602	O76602 caenorhabdi
39	297	6.1	1459	5	O17084	O17084 caenorhabdi
40	289	5.9	1001	3	Q05164	Q05164 saccharomyc
41	287	5.9	716	4	Q9NVE4	Q9nye4 homo sapien
42	286.5	5.9	1321	4	O14594	O14594 homo sapien
43	284	5.8	512	4	Q9H3R2	Q9h3r2 homo sapien
44	282	5.8	513	4	O43418	O43418 homo sapien
45	281	5.8	866	12	O39781	O39781 equine herp

ALIGNMENTS

RESULT 1

Q9ULI3 ID Q9ULI3 PRELIMINARY; PRT; 654 AA.
AC Q9ULI3;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE KIAA1237 PROTEIN (FRAGMENT).
GN KIAA1237.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN.
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K., Kikuno R., Hirose M., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
RT the complete sequences of 100 new cDNA clones from brain which code
RL DNA Res. 6:337-345(1999).
DR EMBL: AB033063; BAA86551.1; -
DR InterPro: IPR000152; Asx_Hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR Pfam: PF00008; EGF; 3.
DR SMART: SM00179; EGF_CA; 1.
DR SMART: SM00001; EGF_Like; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
DR PROSITE: PS01187; EGF_CA; 1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
KW NON_TER
SQ SEQUENCE 654 AA; 70602 MW; 8F51D7F639E8B0CE CRC64;

Query Match

70.2%; Score 3409; DB 4; Length 654;

Best Local Similarity 100.0%; Pred. No. 1.3e-199;	
Matches 654; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 293 SAPLSVSQTTLPQSSSTPVLPRARETPTVTSQTMTSEMTMLHSSQTADLKSTQHPQE 352	
DB 1 SAPLSVSQTTLPQSSSTPVLPRARETPTVTSQTMTSEMTMLHSSQTADLKSTQHPQE 60	
QY 353 KVITESKSPSLVSLPTSTKAVTNSPLPPSLTESSTEQTLPATSTNLAQMSPTFTTIL 412	
DB 61 KVITESKSPSLVSLPTSTKAVTNSPLPPSLTESSTEQTLPATSTNLAQMSPTFTTIL 120	
QY 413 KTSOPLMTTPTGLSTASLVTPIAVOTTAGKQLSLHPEILVPOISTEGGISTERNRVI 472	
DB 121 KTSOPLMTTPTGLSTASLVTPIAVOTTAGKQLSLHPEILVPOISTEGGISTERNRVI 180	
QY 473 VDATTGLIPTSVPTS KAKEMTKLGVTAEYSPASRSISLGTSPSPOTTVVSTAEDLAPKSAT 532	
DB 181 VDATTGLIPTSVPTS KAKEMTKLGVTAEYSPASRSISLGTSPSPOTTVVSTAEDLAPKSAT 240	
QY 533 FAVQSSSTQSPSTLSSASVNSCAVNPCLHNGECVADNTSRGYHCRCPPSWQGDGDCSYDVN 592	
DB 241 FAVQSSSTQSPSTLSSASVNSCAVNPCLHNGECVADNTSRGYHCRCPPSWQGDGDCSYDVN 300	
QY 593 ECLSNPCPSTATCMTNTOGSCFICKCPVGQLEKGCICNLVRTFVTEFKLKRFLNTTVEKHS 652	
DB 301 ECLSNPCPSTATCMTNTOGSCFICKCPVGQLEKGCICNLVRTFVTEFKLKRFLNTTVEKHS 360	
QY 653 DLQVEVEIETKLNCFPSALPSYIRSVVHASRESNAVVISLQTFSLASNVTLFDLADM 712	
DB 361 DLQVEVEIETKLNCFPSALPSYIRSVVHASRESNAVVISLQTFSLASNVTLFDLADM 420	
QY 713 QKVNCKSSAEVQCLGSGRRIFRAGSLCKRKSPCEDKDTISICTDLGVALCOCKSGYF 772	
DB 421 QKVNCKSSAEVQCLGSGRRIFRAGSLCKRKSPCEDKDTISICTDLGVALCOCKSGYF 480	
QY 773 QFNKMDHSCRACEDGYRLENETCMSCFPGLGGLNCGNPYQLITVVIAGGGLLLIIGIA 832	
DB 481 QFNKMDHSCRACEDGYRLENETCMSCFPGLGGLNCGNPYQLITVVIAGGGLLLIIGIA 540	
QY 833 LIVTCCRKNNKNDISKLFKSGDFOMSPYAEYKPNRSGQEWGREAIEMHENGSTKNLQMT 892	
DB 541 LIVTCCRKNNKNDISKLFKSGDFOMSPYAEYKPNRSGQEWGREAIEMHENGSTKNLQMT 600	
QY 893 DVYYSPTSVRNPELERNGLYPAYTGLPGSRHSCIFPGQYNPSFISDSRRDYF 946	
DB 601 DVYYSPTSVRNPELERNGLYPAYTGLPGSRHSCIFPGQYNPSFISDSRRDYF 654	
RESULT 2	
Q9BSVO PRELIMINARY; PRT; 245 AA.	
AC Q9BSVO;	
DT 01-JUN-2001 (TrEMBLrel. 17, Created)	
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)	
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)	
DE HYPOTHETICAL 27.3 KDA PROTEIN (FRAGMENT).	
OS Homo sapiens (human)	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
OX NCBI_TaxID=9606;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC TISSUE-OVARY, ADENOCARCINOMA;	
RA Strausberg R.;	
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.	
DR EMBL; BC004539; AAH04539.1; -.	
KW Hypothetical protein.	
FT NON_TER 1	
SQ SEQUENCE 245 AA; 27334 MW; OCCAFBE55822EF47 CRC64;	

Query Match 27.1%; Score 1315; DB 4; Length 245;

Best Local Similarity 100.0%; Pred. No. 1.1e-72;

		Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	705	LFDLADRMQKVCNCKSSAEVQCLGSGRRIFRAGSLCKRKSPCEDKDTISICTDLGVAL	764
Db	4	LFDLADRMQKVCNCKSSAEVQCLGSGRRIFRAGSLCKRKSPCEDKDTISICTDLGVAL	63
Qy	765	COCKSGYGFQFNKMDHSCRACEDGYRLENETCMSCFPGLGGLNCGNPYQLITVVIAGGG	824
Db	64	COCKSGYGFQFNKMDHSCRACEDGYRLENETCMSCFPGLGGLNCGNPYQLITVVIAGGG	123
Qy	825	LLLILGIALIVTCCRKNNKNDISKLFKSGDFOMSPYAEYKPNRSGQEWGREAIEMHENG	884
Db	124	LLLILGIALIVTCCRKNNKNDISKLFKSGDFOMSPYAEYKPNRSGQEWGREAIEMHENG	183
Qy	885	TKNLQMTDVYYSPTSVRNPELERNGLYPAYTGLPGSRHSCIFPGQYNPSFISDSRRD	944
Db	184	TKNLQMTDVYYSPTSVRNPELERNGLYPAYTGLPGSRHSCIFPGQYNPSFISDSRRD	243
Qy	945	YF 946	
Db	244	YF 245	
RESULT 3			
Q23587	PRELIMINARY; PRT; 3507 AA.		
ID	Q23587		
AC	Q23587		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	2K783.1 PROTEIN.		
GN	2K783.1.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BRISTOL N2;		
RX	MEDLINE=941150718; PubMed=7906398;		
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,		
RA	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,		
RA	Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,		
RA	Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,		
RA	Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,		
RA	Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,		
RA	Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,		
RA	Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,		
RA	Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,		
RA	Watson A., Weinstein L., Wilkinson-Sproat J., Wohlman P.;		
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.		
RT	elegans".		
RL	Nature 368:32-38(1994).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BRISTOL N2;		
RA	Favello A., Vaudin M.;		
RT	"The sequence of C. elegans cosmid 2K783.";		
RL	Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BRISTOL N2;		
RA	Waterston R.;		
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; U13646; AAC24418.1; -.		
DR	HSSP; P00736; IAPQ.		
DR	InterPro; IPR000152; Asx_hydroxyl.		
DR	InterPro; IPR000561; EGF-like.		
DR	InterPro; IPR001881; EGF_Ca.		
DR	InterPro; IPR001254; Trypsin.		
DR	InterPro; IPR001507; zona_pellucida.		
DR	Pfam; PF00008; EGF; 15.		
DR	SMART; SM00179; EGF_CA; 12.		

[illegible]

Query Match		8.0%;	Score 387.5;	DB 5;	Length 2232;
Best Local Similarity		22.8%;	Pred. No. 5.2e-15;		
Matches 233;		Conservative 145;	Mismatches 395;	Indels 251;	Gaps 34;
QY	4	TETVSRVAPM---RGGEITAHMLLTNSTTADVTGSSASYPEGVNASVLTFQSDSTVQS	60		
Db	1276	TPNFSQSTSPVSVTTTGEMTSH---GSTQTPSTIGTVPQSTVSGS---NSSGSGVTI	1328		
QY	61	GGSHALGDRSYSES-----SSTSS-----SESL	84		
Db	1329	GSSEASTSGSFKTSPSSISPVPPTSPIPTFRASSTGSGTISDVSVSVTSLAPLSSSL	1388		
QY	85	NSSAPRGRSTAGISYQVGRGA---IQRTSSDHTD-----HTYLSSTFT-KGERALL	134		
Db	1389	PSTVPSTQSFSTSESGSKASPVPSQTSPTNTPTGTSTESSTLLSSITSGSTQHTM	1448		
QY	135	SITDNSSSDIVESST--SYKINSNSHSEYSSFSHAQTE---RNSISSYDGEYAQPSTES	190		
Db	1449	SKASSGSTSPSTNSQTGSTVTMGSSSTGCVSTSSASSTQPMSTSQGSSAGSTVASSTAS	1508		
QY	191	PVLHTSNLPSVPTPINPNTSVVLDTDAEFVSDSS-----SSSSSSSSSSSSG	238		
Db	1509	PAA--SSTAPSTGMTSSTSGTSGTISESSTTASASSQTGSTVTMGSSSTSGVSTSSAS	1567		
QY	239	PPLPLPSVQSHHLFSSILPSTRASVHLKSTSDASTPWSSP---SPLPVSLLTSTSA	294		
Db	1568	STQPMSTSQGSSAGSTVASSTAGLVSTVSPSTGMTSGTSSCTVGTSGTISESSTTASAS	1627		
QY	295	PLSVSQTLTPOSSSTPVLPRARETPTVTSFQTSMTSFTMTLHSSQTADLKSQSTPHQKV	354		
Db	1628	SQTGSTVTMGSSSTGCVSTSSASSTQPMSTSQGSSAGSTVASSTTGLVSTVTPSSTGT	1687		
QY	355	ITESKPSLVLPSTKAVTNSPLPPLSTESTETLPTATST-----NLAQMSPT	406		
Db	1688	MGSTSSGTGVTGISTESSTAAASQTGSTVTMGSSSTGCVSTSSAGSQPMSTSQGSSA	1747		
QY	407	FTTILKTSQPL--MTTP---GTLSSSTASLVTC-----PIAVQTTA--GKQLSLTHPEI	453		
Db	1748	GSTVVSSTASPAASSTAPSTGTMSTSSSTSGTGTMSQSTAASTTHTSGTIVTLGSSST	1807		
QY	454	LVPOISTEGGISTERNRVIDATTGLPLTSVPPTSAREM--TTKLG-----VTAEY	502		
Db	1808	SSNMSTSQGSSV--GSTVASSTAGLVSTVSPSTGMTSGTSSCTVGTSGTISESSTASA	1865		
QY	503	SPASRSLGTSPTQTTVVST-----AEDLAPKATFAVQSGST- 539			
Db	1866	SSQTGSTVTMGSSSTGCVSTSSASSTQPMSTSQGSSAGSTVASSTAGLVSTVTPSSTG	1925		
QY	540	---QSPPTLSSASVNSCAVNPCLNHGECVADNTSRGYHCRCP-----PSWQGGD	586		
Db	1926	TMGSTSGTGTGVTISESSTAASSTASQTGSTVTIGTSGTNPSPRSLSQITITPSPSQST	1985		
QY	587	CSVDVNECLSNPCPSTATCNTQGSFCKCPGVQLEKGCINLVRFVTFEKLKRTFLNT	646		
Db	1986	ESTQTSPLPSSSPSTSHVSSEGT-----TMSSGATSGDKMSFLSS 2028			
QY	647	TVEKHSDLQEVENEITKLNCFESALPSYIRSTVHVASRESNAVVISLQTTESLASNVTLF	706		
Db	2029	T-----GTVVSFSSRGS-----SLATTSAPKPSVTCL 2055			
QY	707	DLADRMQK---CVNSCKSSAEVCOLLGSGRRI-----FRAGSLCKRKSPEC 749			
Db	2056	FMYDQTSKIDQTAINTKYTFENFALLVASKLNNESTLGTIDNFGYSAGLNHDQHYPTD	2115		
QY	750	D-----KDTISICTDLGDVALCOCKSGYQFQFNKMDHSCRACEDGYLENETCM	796		
Db	2116	DYNGIKSVPPFDGTDGDDLDLKD---DKSLATADWTTPVADQTCM 2160			
QY	797	----SCPFQ--LGGNLGCONPYQLITVWTAAGGGLLLLTGTA----LIVTCRKNKNDISK	847		
Db	2161	IFISAPEDEYGGTTIKSTTYFETVGVVLGGAKSIPGLSIDKNIVITNTNMNDRASA	2220		

QY	848	LIFK 851				
Db	2221	VWSK 2224				
RESULT	5					
Q99QY4	PRELIMINARY;	PRT;	2271	AA.		
ID	O99QY4					
AC	O99QY4;					
DT	01-JUN-2001 (TrEMBLrel. 17, Created)					
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)					
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)					
DE	SA2447 PROTEIN (HYPOTHETICAL PROTEIN SAV2654).					
GN	SA2447 OR SAV2654.					
OS	Staphylococcus aureus (strain N315), and					
OS	Staphylococcus aureus (strain Mu50).					
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;					
OC	Bacillus/Staphylococcus group; Staphylococcus.					
OX	NCBI_TaxID=158879; 158878;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);					
RX	MEDLINE=21311952; PubMed=11418146;					
RA	Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,					
RA	Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,					
RA	Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,					
RA	Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,					
RA	Sekinizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,					
RA	Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,					
RA	Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;					
RT	"Whole genome sequencing of methicillin-resistant Staphylococcus					
RT	aureus.";					
RL	Lancet 357:1225-1240(2001).					
DR	EMBL; AP003138; BAB43752.1; -					
DR	EMBL; AP003366; BAB58816.1; -					
DR	InterPro: IPR001899; Gram_pos_anchor.					
DR	Pfam: PF00746; Gram_pos_anchor; 1.					
DR	PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.					
KW	Complete proteome; Hypothetical protein.					
SQ	SEQUENCE 2271 AA; 227844 MW; 7C2A7040D6C8289D CRC64;					
Query Match 7.5%; Score 364; DB 16; Length 2271;						
Best Local Similarity 22.6%; Pred. No. 1.4e-13;						
Matches 183; Conservative 151; Mismatches 363; Indels 114; Gaps 22;						
QY	2	SQETVSRVAPMRGGETTAHMLTNSTTADV-----TGSSASYPEGVNASVLTFQS	54			
Db	1054	SDSKSLSTSQSGSTS-----TSTTSASVRTSESQSTSGSMSASQSDMSISTFS	1108			
QY	55	DSTVQSGSHALGDRSYSESSTSSSESLNS-----SAPRGERSIAGISYQVGRGTAIEQ	110			
Db	1109	DSTSDSKASASTA--SSESISQASSTSGSVSTSTSLSTNSERTSTSVSDSTSLSTSESD	1167			
QY	111	RTSSDHPDHYLYLSTFTFKGERALLSTDNSSSDI--VESSTSYTKIS---NSSHSEYSVF	166			
Db	1168	SISESTSDSISIAISASESTSISLESNSTSDSESQSAFSLSESLSESTSESTSES	1227			
QY	167	SHAQTERNSLTSSVDGEVAQPT-----ESPVLHTSNLPSYPTTINMP--NTSVVLDT	216			
Db	1228	SSSTSESTSLSDSTSESGSTSTSLNSTSGSASISTSTISESTSTFKSESVSTSLSMST	1287			
QY	217	DAEFVSDSSSSSSSSSSSGGPPPLPSPVQSQSHLPSLLPSTRASVHLKSTSDASTP	276			
Db	1288	STSLNSTSLSTSLSDSTSDSKSDSLSTSMSTSDSISTSKSDSISTSTSLSGSTSESED	1347			
QY	277	WSSS-----PSPLPVSLTSTTSAPLSVS--QTLLPQSSSTPVLPRAR-----ETPVTSE	323			
Db	1348	STSSSEKSDSTSMISMSQSTSGSTSTSTSLSDSTSTSLSLSASMNQSGVDNSASQ	1407			
QY	324	QTSMTWSFMTLHSSQ--TADLKSGSTPHQKVITES--KPSPLVSLPTESTKAVTTNSPL	380			
Db	1408	SASNSTSTSESDSQSTSTVTTSQSTSQSESTSTSTSLSDSTSIKSTSGSGSTSAAL	1467			

Oy	381	-----PPSLSTSEQTLLPATSTNLQAMSPFTTTTILKTSOPLMTPTGTLSSASTASLV	432
Dd	1468	SGSESDSQSISSTASESTSESASTSLSDSTSTSGNSGSTSTSLSNSASASEDSST	1527
Oy	433	TGPFAVOTAGKOLSLTHPEILLVPOTIEGGISTERNRVIVDATTGLIPLTVSPTSAKEM	492
Dd	1528	S--LSDSTSAMOSSRDSQSSTASLSDSLSTSTS-NRMSTIASLSTSVSTSESGSTSES	1584
Oy	493	TTKLGVTF----ABYSPASRSLSCTSPQTVVYSTAEDLPAPSAFOSSTQSPITLSSS	548
Dd	1585	TSEDSTSTSLSDSQSTSRSTASGSASTSTSTDSTRSTASSTSMRTSTSDSQSMSLS	1644
Oy	549	ASVNSCAVNCLHNGECVADNTRSGYHCRCPPSWGOGDCSYDVNECLS-NPCPSTATCNN	607
Dd	1645	TS-TSTMSNSTSLSDSVSDSTS-----DSTSASTSGMSVSISLSDSTSTS	1690
Oy	608	TGGSFICKCPVGQLKGICNLVRFTVEFKLRKLTFLNTTVEKHSDLQEVENEITKTLM	667
Dd	1691	TSASEVNSASI-----SDSQMSSESVNDESEV	1717
Oy	668	CFSALPYIRISTVHASRENAVVISLQTTFF---SLASNVTIFDLADMQRKVNSCKKSA	723
Dd	1718	SENSESDSKMSGSTSVSDGSLSVSTSLRKSESVSESSLGSLGSDSVSTSDSSSL	1777
Oy	724	EVCOLLGSRRIFRAGSLCKRKSPCEDKDTS	754
Dd	1778	SVTSLRSSVESDESLSDSKSTSGSTSTS	1808
RESULT	6		
ID	014651	PRELIMINARY;	PRT; 957 AA.
AC	014651;	014650;	
DT	01-JAN-1998	(TEMBLrel. 05, Created)	
DT	01-NOV-1998	(TEMBLrel. 08, Last sequence update)	
DT	01-JUN-2001	(TEMBLrel. 17, Last annotation update)	
DE	MUCIN (INTESTINAL MUCIN) (FRAGMENT).		
GN	MUC3.		
OS	Homo sapiens (human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	GMEJNE-97476275; PubMed-9334251;		
RA	GUM J.R. Jr., Ho J.J.L., Pratt W.S., Hicks J.W., Hill A.S.,		
RA	Vinall L.E., Robertson A.M., Swallow D.M., Kim Y.S.;		
RT	"MUC3 human intestinal mucin. Analysis of gene structure, the carboxyl		
RT	terminus, and a novel upstream repetitive region.";		
RL	J. Biol. Chem. 272:26678-26686(1997).		
DR	EMBL; AF0071194; AAC02272.1; -		
DR	EMBL; AF0071196; AAB84383.1; -		
DR	EMBL; AF0071195; AAB84382.1; -		
DR	InterPro; IPR000561; EGF-like.		
DR	SMART; SM00181; EGF; 1.		
DR	PROSITE; PS00022; EGF_1; UNKNOWN_2.		
DR	PROSITE; PS01186; EGF_2; 1.		
KW	EGF-like domain; Glycoprotein.		
FT	NON_TER	1	
SQ	SEQUENCE	957 AA; 99554 MW; E6772256D600A1F CRC64;	

		Query Match	7.5%; Score 363.5; DB 4; Length 957;
		Best Local Similarity	24.2%; Pred. No. 5.2e-14;
		Matches 240; Conservative	124; Mismatches 333; Indels 295; Gaps 49;
Qy	1	MSQETVSR-----VAPRGGEITAHWL--LTNSTSDVTGSSASYPEGVNASVLTFQ-	53
		: : : : : : : : : : : : : : : : : :	
Dd	52	ITTTTSHSTPTSPSTRRTETVSSTPSTSNNTTETT--SHSTPSYITSITTTP	109
		: : : : : : : : : : : : : : : : : :	
Qy	54	SDSTVQSGSHTALGDRSYSE---SSSTSSSELNSSAPRGERSIAGISVGQVRGTAEIQ	110
		: : : : : : : : : : : : : : : : : :	

Db	110	SSSTPSPSSSI	TTTTTSHSTPGTSSITTTTSHSTPSTSSII	-----	154		
Qy	111	RTSSDHTDHTYLS	FTFKGERALLSI	TDNSSSSDIVESSTSYI	----KINSNSHSEYSSF	166	
Db	112	TTTTTTSHD	-TSPFS	-----	167		
Db	155	TTTETTS	SHD	-TSPFS	-----	205	
Qy	167	SHAOTERNIS	SYDGEYAQPSTES	PVLHNTNLPSYTPPINMPT	---SVVLDTDAEFSD	223	
Db	206	TTTETTS	SHARSFTSSIT	TTETTS	-----	260	
Qy	224	SSSSSSSSSS	SSSGPPLP	-----	273		
Db	261	TTSHSTP	SSSI	TTTTETPLHSTPG	LSWVTTTKTSHITPG	320	
Qy	274	-----	STPSSSSP	PLSVLTST	TSAPLSVSQ	319	
Db	321	TSSIT	TTETSETP	-SLSST	IVSTVSTYALTSHFTSE	379	
Qy	320	VTSQT	STMTS	FMTHLSS	QADLKSO	ST-----	370
Db	380	TTSLR	LTPSP	SVGTSL	TTTTDP	PSIPTD	437
Qy	371	-----	TKAVT	NSPL	-----	398	
Db	438	TSPTM	STVRMT	LRITEN	TISSFS	TSIVWIPETQ	497
Qy	399	-----	NLAOMSP	-TFMT	ILKTSOPLMT	PGTLLS	451
Db	498	TDS	STLHLTP	STALST	IVSTSQ	--VP	549
Qy	452	EILV	PQISTEG	GI	STERNV	-----	505
Db	550	SEFT	TESFTR	-ST	SNAIL	SPSTII	600
Qy	506	SRS	LGTS	SPSP	QTTVV	STAE	564
Db	601	IHS	VPSSP	-----	YIFST	ENVGS	612
Qy	565	CVAD	NTSR	GYHCR	CP	-----	612
Db	656	SLP	STP	-----	CPG	ITIV	705
Qy	613	ICK	CPV	QLEK	ICNL	VR	661
Db	706	MV	CTP	TSIS	IQ	-----	747
Qy	662	TK	L	NMCF	-----	-----	721
Db	748	TVPT	NTVFT	STR	LPTSET	WL	807
Qy	684	---	RESNA	VIS	LQ	TF	721
Db	808	PP	TR	SET	PA	VT	867
Qy	722	---	SAE	V	COL	LS	779
Db	868	G	F	S	D	R	900
Qy	780	SCR	ACED	GY	R	LEN	810
Db	901	SCE	FA	V	EQ	V	930
RESULT	7						
Q9UKW9	ID	Q9UKW9	PRELIMINARY;	PRT;	1217	AA.	
AC	Q9UKW9;						
DT	01-MAY-2000	(Tremblrel. 13, Created)					
DT	01-MAY-2000	(Tremblrel. 13, Last sequence update)					
DT	01-DEC-2001	(Tremblrel. 19, Last annotation update)					
DE	INTESTINAL	MUCIN 3 (FRAGMENT).					
GN	MUC3.						
OS	Homo sapiens	(Human).					

OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OX	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC	NCBI_TaxID:9606;
RP	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=INTESTINE;
RC	MEDLINE=99443732; PubMed=10512748;
EX	Crawley S.C., Gum J.R. Jr., Hicks J.W., Pratt W.S., Aubert J.P.,
RA	Swallow D.M., Kim Y.S.;
RT	"Genomic organization and structure of the 3' region of human MUC3:
RT	alternative splicing predicts membrane-bound and soluble forms of the
RT	mucin.";
RL	Biochem. Biophys. Res. Commun. 263:728-736(1999).
DR	DR EMBL; AF113616; AAF13032.1; -.
DR	DR InterPro; IPR000561; EGF-like.
DR	DR InterPro; IPR000082; SEA.
DR	DR Pfam; PF01390; SEA; 1.
DR	DR SMART; SM00181; EGF; 2.
DR	DR SMART; SM00200; SEA; 1.
DR	DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR	DR PROSITE; PS01186; EGF_2; 1.
DR	DR PROSITE; PS50024; SEA; 1.
KW	EGF-like domain; Glycoprotein.
FT	NON_TER 1
FT	SEQUENCE 1217 AA; 129171 MW; 0783AA4003C969664 CRC64;

Query Match 7.5%; Score 363.5; DB 4; Length 1217;
Best Local Similarity 21.9%; Pred. No. 7e-14;
Matches 231; Conservative 138; Mismatches 374; Indels 311; Gaps 44;

Qy	1	MSQETVSRSVAPMRGEITAHMLLNSTISADVTGSSA-SPEGVNVAVLTFQSDSTVQ	59
Db	188	ITTKTTHSHSTPS-----FTSSIIITTTTSHSARSSTSIITTTT--SHNTRS	234
Qy	60	SGSGHTALGRDYSSESSTSS---SESLSNAAAPGRSAGISYGQVGRATGEOR----	111
Db	235	FTSSIIITTEINSHSTISFTSSIIITTTTSHSTSPFSSSIT-----TTETPLHSTPGLTS	288
Qy	112	--TSSDHTD-----TVLSSTFKGERALLSITDNSS-SSDIVESSTSYIKIKN	157
Db	289	WVTTTKTTHITPCLSSIIITTTTSHSTPGFTSSIIITTTTSESPLSSSIIYSTVST	348
Qy	158	SSHSEYSFSHAGTQSRNISYSOGYEAQPTSESPVLHSLNLSYPTINMPTNVSVLDRD	217
Db	349	STRAITSHFTTSET-----AVTPTPVTFSSLSLTDIPTTSLRSLTSPSSVGTSTSLTTTD	402
Qy	218	AEFVDSOSSSSSSSSSSSSSGPPLPLPSVYQSHHLSFSSILPSTRAVHLKTSDDAS---	274
Db	403	FPSIPTDILPTRTHIISSPSIQSTETSSSLVGTTSPTMTSVRMTLRITENPTISSEST	462
Qy	275	-----TPWSSSP-----SPLVSLT-----TSTSAPLSVSQ	300
Db	463	SIVVIPETPTQTPVLTSATGCTQTSAPYIVTFGSDSSLSLHLTLPSTALSTIVSTSQ	522
Qy	301	TTLPQSSS-----TPVLPRARETPTVTSF-----QTST---MTSFMTLHSSQFA	341
Db	523	VPISPSTHSTLQTTPTSP-SLQTSLSATSEFTPESTFRGSTSNALLTSFSTIINSSTPT	581
Qy	342	DLKSQS-----TPHQEKVI-----TES-KPSPLSVSLPTTESTKAVTNSPLPSP	383
Db	582	IIMSSSPSSASIIIPVSTTHSVSPSPYIFSTENVGASLITGPPSLUSSATISTSTSSS	641
Qy	384	LTESSTEQTLPATSTNLAQMSPP--TFTTUIL-----KTSQPLMTPTGT	424
Db	642	LTTALTLEIT-PFYSILSPSTTPCGTITITIVPASPTDPCVMDPSTEATSP--TTPLT	698
Qy	425	L-SSTASLVNCP--IAQVTTAGQL--LUTHPE--ILVPQISTEGGISTERNRVIVDAT	476
Db	699	VFPFTEWVTCPSISIQTLTLYMTDSSMMPESSSISPNASSSTGCTGVPTNTVFTST	758
Qy	477	-----TGLPIPLTSYPTSAKEMTTLKGVTAEYSPASRSLSLGTSPSPQTTVVSTAD	525

Db	759	RLPSETWLNSSVILPLPLPGVSTIPLTWKPPSSSLPTILRTSSKTHSPPTTRTSETP-	811
Qy	526	LAPKSATFAVQSSTQSPPTLSS-----SASVNSCAVNPCLHNGECVADNTSRG	573
Db	818	-----VATTQPTTLTSRRTRRITTSQMTTQSTLTFTA-----GTCNDGGTWEQ	860
Qy	574	YHCRPPSWGDDCSDVDVNECLSNPCPSTATCNN--TQGSFTCKCPGVQLEKGCINLVR	631
Db	861	GQCACLFPGFSGDRCOLQTR-----CONGGQMDGLKQCP-----	894
Qy	632	TFVTEFLKLTFLNTTVE---KHSDLQEVENETTKLNNCFSSALPSYIRSTVHASRESNA	688
Db	895	-----STFGSSCEFAVEQVDLDVTEVTEGVMSVDQOFSPLNDNTNSQAVRDENK	945
Qy	689	V-----VISLQTTFSLASNVTLFD-----LAD	710
Db	946	TFWNQMKIFADMQGFTEFKGVELTSLRNGSIVVDYLVLLEMPFSPQLESEYEQVKTTLKE	1005
Qy	711	RMQKC---VNSCKSSAEVC-----QLLGSQRRIFRAGSLCKRKSPCEDKDTSI----	757
Db	1006	GLQNASODVNSCQDSQTLCKFKPDSIKVNNNSKTELTPAAICRAAPTGYEEFYFPLVEAT	1055
Qy	758	DLQGVALC-----QCKSGYGFQFNKMDHSCRA-CEDGYRLENETCMSCPFG-----	805
Db	1066	RLRCVTKCTSGVDNADICHQGCQVLTSGPTCRCTSTDTWHFSGPRCEVAVHWRALVGGL	1125
Qy	806	NCGNPYOLITVIAAAGGGLLLILGIALIVTCOR	839
Db	1126	-----TAGAALLVELLLALGVRAVR	1145
RESULT 8			
ID	Q9C105	PRELIMINARY;	PRT; 1236 AA.
AC	Q9C105;		
DT	01-JUN-2001 (T-EMBLrel. 17, Created)		
DT	01-JUN-2001 (T-EMBLrel. 17, Last sequence update)		
DE	01-DEC-2001 (T-EMBLrel. 19, Last annotation update)		
DE	PUTATIVE GLUCOAMYLASE I (ALPHA-1,4-GLUCAN GLUCOSIDASE), EXTRACELLULAR		
DE	SPARCH-DEGRADING ENZYME, BY SIMILARITY TO S. CEREVISIAE STAL, CONTAINS		
DE	CHITINASE FAMILY SIGNATURE.		
GN	SPAB1E7.04C.		
OS	Schizosaccharomyces pombe (Fission yeast).		
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;		
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;		
OC	Schizosaccharomycetes.		
NCBI_TaxID=4896;			
EN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=972H-;		
RA	Wood V., Rajadream M.A., Barrell B.G., Seeger K., Harris D.;		
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AL590605; CAC36921.1;		
DR	InterPro; IPR001917; AminoTransf_2.		
DR	InterPro; IPR001579; Chitinase_2.		
DR	Pfam; PF00192; chitinase_2.		
DR	PROSITE; PS00599; AA_TRANSFER_CLASS_2; UNKNOWN.1.		
SO	SEQUENCE 1236 AA; 123387 MW; 5A2D33A30B87CDD8 CRC64.		

Query Match	7.3%	Score 352.5;	DB 3;	Length 1236;
Best Local Similarity	24.6%	Pred. No. 3.3e-13;		
Matches 206;	Conservative 132;	Mismatches 361;	Indels 139;	Gaps 27;

[illegible]

```
Db 498 STSSEVSEVSTLLSGSAIPSPSSSSPSSSISSPMTSV-LSSSSSIPTSSSDSSS 556
Qy 173 RSNISSYGEYAQPSTSPVLHNLPSYPTINNPNTSVVLDTAEFVSDSSSSSSSS 232
Db 557 ITTSSGIGSSIPSTSSV--SSILSSSTSPSSSTLSI-----SSSTSTSTFS 604
Qy 233 SSSSGPLPLPSVQSQHLLPSTRASVHLKST--SDATPSSSSSPPLPVSLTT 290
Db 605 SASTSSSSSSSSSSSTLLSSPTPT--SSLMTSSSIISGSSILSSITPISSSL 663
Qy 291 ST-----SAPLSVSQTLPLQSSSTPVLPRARETPVTFQSTMT--SEMTM 334
Db 664 STYSSVIPSSSTLVSSSSSIVSSPVASSSSPI-----PSSSLVSTYASLSNI 716
Qy 335 LHS--SOTADLKOSTP---HOEKVITESPVLVSTESTKAVTT-----NSPLPPS 383
Db 717 THSSLTAMSSSAIPTSVNSSTLITASSNTLLSSITSSSAIVSVSTVNISSNLP 776
Qy 384 -----LTESSTQTLPATSTNLAQSPFTTILKTS--QPLMTPTGTLSSLASLVG 434
Db 777 TASSQQLNTSLATSLYLSSSSSRTISSTNEYNTSFHAPTSSVSTSSSTSLAAN 836
Qy 435 PIAYQTAGKOLSTHPEILVPLQISTEGGISTERNRIVDATGLI---PLTSVPTSAKE 491
Db 837 K-GVNSISITLNLE-----STSSVTSTAYTTDSVTSTALTSGQSSSVSSLS 886
Qy 492 MTKLG-----VTAYSPASRLGSPQTTVVSTAEDLAPKATFAVQSTQ----- 540
Db 887 STTSLSTIPVTSSVAPAVTSTGETS---SVGSGTDSATSSSWTAETSSAITSSVA 943
Qy 541 --SPTTLSSASVNSCA-VNPLCHNGECVADNTSRGYHCRPPSMQGDGCDVDVNECLN 597
Db 944 SVTFTSSSSASSSSSEVDP-----STAASATGSSSTSSATASVSGSSSTSVATASATD 998
Qy 598 PCPSTATCNGTQFCICKCPVGYOLEKICNLVFTFVEPKLRTFLNTVVEKHSDLQEV 657
Db 999 SSTSSIAAASVTGSSSVATASVATDSSSTSSVATASATD-----SSTSSIAVASVTS 1051
Qy 658 ENEITKLNCFSPALPSVIRSTVHASRESNAVVISLQTTFLSLANVILFDLADRMOKCVN 717
Db 1052 STSSVATASATDSSSVATASVATISGSSSIATASV--TGSPTSSVAVSTSSVEGTAS 1109
Qy 718 SCKSAEVCQLGQRRIFRAGSLCKRSPCEDKDTISCTDGLDVALCQCKSGYFOFN 775
Db 1110 STIAAASATLSSDA---ASGSSTVSSATSSSSSNAATADSSVTDTSPSDFNAN 1164

RESULT 9
Q9N4S7
ID Q9N4S7 PRELIMINARY; PRT; 1079 AA.
AC Q9N4S7
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Y51B11A.1 PROTEIN.
GN Y51B11A.1
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Cotton M.;
```

```
RT "The sequence of C. elegans cosmid Y51B11A.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006797; AF60743.1; -.
DR InterPro; IPR002965; P_Rich_extensn.
DR PRINTS; PRO1217; PRICHEXTENS.
SQ SEQUENCE 1079 AA; 110532 MW; 8DBDE3824CF80CA1 CRC64;
```

```
Query Match 7.28; Score 348; DB 5; Length 1079;
Best Local Similarity 24.5%; Pred. No. 5.3e-13;
Matches 176; Conservative 102; Mismatches 324; Indels 116; Gaps 26;

Qy 2 SOTETVSRSVAPMGGEGITAHWLL-----TNSTTTSADVTTGSSASYPEGVNASVLTFQSDS 56
Db 291 TSTPEPSSSTTPVQTTTITAPETTSTPEPSSSTTPVQTTTITA--PETTSTPEPSS--STT 347
Qy 57 TVQSGGSHALGDRSYSESSSTSSSESLNSAPR---GERSIAGISYQVGRGTAEIQRTS 113
Db 348 PVQTTTITAPETTTEPSSSTTPVQTTTITAPETTSTPEPSSSTTPVQTTTITAPETTS 407
Qy 114 SDHTDHT---YLSSTFKGERALLSITDNSSSDIVESSTSYIKISNSHSEYSSFSHAQ 170
Db 408 TEPPSSSTTPVQTTTITAPETT--STEPSSSTTPVQTTTITAPETTSTPEPSSSTTPVQ 465
Qy 171 TERNISSYDGEYAQPSTSPVLHNLPSYPTINNPNTSVW-----LDTDAEFVSDS 225
Db 466 TTTTITAPETT-----TTSTPEP--SSSTTPVQTTTITAPETTSTPEPSSSTTPVQTTTITA 517
Qy 226 SSSSSSSSSSSSGGPL-----PLPSVSQSHLFPSSILPSTRASVHLKSTSDASTPWSSS 280
Db 518 PETTSSPSSSTTPVQTTTITAPETTSTPEPSSSTTPVQTTTITAPETTSTPEPSSSNTPVQ 575
Qy 281 PSLPLVSLTT-----STSAPLSVSOTTLPO-----SSSTPVL----- 312
Db 576 TTPVQTTTITAPETTSTPEPSSSTTPVQTTTITAPETTSTPEPSSSTTPVQTTTITAPET 635
Qy 313 -----PRARETPVTSFQSTMTSFMTLHSSQTLADLKSQSTPHQEKVITESKPSLVSLP 367
Db 636 TSTEPSSSTTPVQTTTITAPETTSTPEPSSSTTPVQTTTITAPETTSTPEPSSSNTPVQ 695
Qy 368 TESTKAVTNSPLPPLSTES---STEQTLPATSN---LAQMSPTFTTIL----- 412
Db 696 TTTTITAPETTSTPEPSSSTTPVQTTTITAPETTSTPEPSSSTTPVQTTTITAPETTSTEP 755
Qy 413 --KTSQPLMTTPGTLSSSTASL-----VTGPIAVOTTAGKQLSLTHPEILVPOISTEGGIS 465
Db 756 PSSSTTPVQTTTITAPETTSTPEPSSSNTPVQTTTITAPETTSTEP-----PSSSTS---P 808
Qy 466 TERNRIVDATATGLIPLTSVPTSAKEMTKLGVTAEYSPASRSLGTSPSPQTTVVSTAED 525
Db 809 VQTTTITAPETTSTPEPSSSNTPVQTTTITAPETTSTPEPSSSTTPVQTTTITAPETTST 868
Qy 526 LAPKSAFAVQSSST-QSPSTLSSASVNSCAVNCPLHNGECVADNTSRGYHCRCPSPSQG 584
Db 869 EPPSSSTTPVQTTTITAPETTSTPEP--SSSTTPVQTTTITAPETTSTPE-----PPSSST 921
Qy 585 --DDCSVDVNECLSNPCPSTAT--CNNTQGSFCKCPVGYOLEKICNLVRTFTVTEF 637
Db 922 TPVQTTTITAPETTSTPEPSSSTTPVQTTTITAPETTSTPEPSSSNTPVQTTTITAPETT 974

RESULT 10
Q29071
ID Q29071 PRELIMINARY; PRT; 528 AA.
AC Q29071;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
```

Qy	10	SVAPMRGEITAHLLNTSTTADYTGSSASYPEGVNASVLTQFSDS-----TVQSGG	62
Db	35	SVQPSSSGAP-----TTSATSVQ-TSSSSPPISTTISVQSSSVPTTSTTSVQPS	88
Qy	63	SHPALGDRSVSESSSTSSSELNNSAPRCERSIAGISGVQRTAIEQRTESDTHDTYL	122
Db	89	SSGAPTRATSVQSSSSSAPISSTT-----SVQPSSSGVPPTTSATSVQSSSSSAPPT	143
Qy	123	SSTFTKGERALLITONSSSDIVESSTSVIKINSHSEYSSFSHAQTERSNISSYDGE	182
Db	144	SAT-----SVQPSSSSPPISSTSVQPSSSSSAPPTTSAT-----	179
Qy	183	YAPQSTES--PVLUHTNLSYPTTINMPNTSVVLDTDAEFVSDSSSS-----SSSSSSSS	236
Db	180	-VQPSSSSPPPI-----SSTVSVQTSSSSVPTTSTTSVQPSSSSVPTTSATSVRS	231
Qy	237	SGPPLPLPSVQSHLFFSILPSTRASVHLLKSTSDASTPWSSSPPLP-----VSLTST	292
Db	232	SSSSTPIPT-----TSVQPSSSSAPPTTSAT-----VQPSSSSPTIPSTTSVQPSSS	282
Qy	293	SAPLSVQSOTTLPOSSSTPVLPARETPTVTSFOTSTMTSFMMLHSSQTLADLKSO	352
Db	283	SAPTSATSVQPSSSSP-----PISS-----TSVQPSSSSSSP	317
Qy	353	KVITESKSPSLV-SLPETSKAVTNPSLPPLSSTSTEQTLPATSTMLAOMSPFTTTI	411
Db	318	TTSTTSVQPSSSGAPPTTSATSVQPSSSSPPPI--SSTISVQPSSSS-----SPTTSTS	371
Qy	412	LKTSQPLMTTPGTLSTASLVTGPFAVQTACKQLSLTHPELVPQISTEGGISTERNRV	471
Db	372	VQPS-----SSGAPPTTSATSVQPSSSS-----VPTTSATSVRSSSSST	412
Qy	472	IVDATTLGLIP--LTVSPT-----SAKEMTKLGVTAEYSPASRSLCTSPSPOTTVST	522
Db	413	PIPTTTSVQPSSSSVPTTTSATSVQTSSSSSTPISTTSVQPSSSSSAPPTTS-ATSQPS	471
Qy	523	AEDLPKSAATFAVQ--SSTQSPPTTSSSASVNSCAVNP	538
Db	472	SSSSPPISTISVQPSSSSSSPTTSTTSVQPSSSSGAP	509

268	QY	KTSDASTPWSSSPPLPVSLTTSTAPLSVSOTTLPQSSSTPVLPRARETPTVTSQTST	327
:	:	: : : : : : : : : : :	:
3	DB	QSSGGTSTPTTATQP---TSTSTQP---GTTQLLSTTSTPTTATQPT---STST	50
328	QY	MTSEMTMLHSOTADLKSQSTPHQEKVITBSKPSLSLP-TSTKAVITNSPLPPLSTE	366
:	:	: : : : : : : : : : :	:
51	DB	QTPTQTQLPSTSTPTTATQP---TXTSTQTPGTQLPGTSTPTTATQPTSTSFQT	106
387	QY	SSTEQTLPATSTNLAQSPSTFTTILKTSOPLMT---TPQTLSTASLVTPGIAVQTAG	443
:	:		:
107	DB	PGTQLPSSST---PTTAT---QPTSTASQTPGT-----	136
444	QY	KQLSLHPEILVPOISTEGGISTERNRIVDATTGLTLPFSVPTSAKEMTKLGVAEYS	503
:	:		:
137	DB	---TQP-----PGASS-----PTTV-----TQPTGSSQTP	161
504	QY	PASRLGTSPTQTVVSTAEADLPKSAFAVOSSQSPPTLLSSASVNCANPCLHNG	563
:	:	: : : : : : : : : :	:
162	DB	GTTPPGGASTPTTVTQ---PTGSSQTSQTTQPPGGASST-----	201
564	QY	ECVADNTSRGYHCRCPSPWQDDCSVDVNECLSNPCPSTCNTQGSFICKCPVGQL-	622
:	:		:
202	DB	--VTSSSSTGSN-----DFCNSNCPKSPASCVKLVDSPCLGLEYYN	243
623	QY	EKGICNLVRTFTVEFKLRFTLNTTVEKHSDLQ-----VENEITKT--	664
:	:	: : : : : : : : : :	:
244	DB	NSSCVKGTTFPGE-----IGMSVNETTDELDKNSVNYOTLHSSVVKFFENTFKTDY	296
665	QY	-----LNMCFPSALPSYIRSTVHASRESNAVVISLQITF-----SLAS	701
297	DB	GQTVILKYSKDSLMS--SRSYMRAATQF--YVSVNMFGENTKEDEESVASVKEAVKTD	353
:	:	: : : : : : : : : :	:
702	QY	NVTFLDLARMQ--KCVNSCKSAEVCQLLGSORRIFRAGSLCKKR-----	745
:	:		:
354	DB	NNVERYQQDRCDYGCY---KGSNVC-----RNLQCTCKPGLERLNPQVPFC	400
746	QY	-SPECDKDTs-----ICTDLGDVALCQCKSGYFQFNKMDHSRACEDGYRLENETCMSCP	799
401	DB	VAPTCSPCAEKKOLCLCKKONGAM--EC-----GCMAGYRKANGKCECP	444

Qy	800	FGLGCLNGCPYQLITVWIAAAGGGLLLIIGIALIIVTCRKNKNDISKLIFKSGDFQ	856
Dd	445	FGYGMCDCKDQFOLILTIVGTIAGAFILILLIVFIYSMRSKNKK-----KSGBEQ	494
RESULT 12			
Q21027		PRELIMINARY;	PRT; 786 AA.
ID	Q21027		
AC	Q21027;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DD	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)		
DE	F59A6.3 PROTEIN.		
GN	F59A6.3.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditoidea;		
OC	Rhabditidae; Peloderinae; Caenorhabditis.		
OX	NCB_I_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BRISTOL N2;		
RX	MEDLINE=99069613; PubMed=9851916;		
None:			
RT	"Genome sequence of the nematode C. elegans: a platform for		
RT	investigating biology. The C. elegans Sequencing Consortium.";		
RL	Science 282:2012-2018(1998).		
[2]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BRISTOL N2;		
RA	Nhan M.;		
RT	"The sequence of C. elegans cosmid F59A6.";		
RT	Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.		
[3]			
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BRISTOL N2;		
RA	Waterston R.;		
RL	Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.		
DR	EMBL; U41994; AAK31523.1; -		
DR	InterPro; IPR000436; Sushi_SCR_CCP.		
DR	SMART; SM00032; CCP; 1.		
SQ	SEQUENCE 786 AA; 79211 MW; 60425E32F083B3CB CRC64;		
 Query Match Best Local Similarity 24.0%; Score 339; DB 5; Length 786; Matches 180; Conservative 105; Mismatches 334; Indels 130; Gaps			
Qy	26	TNSTTSADVTCSSASYPEGVNASVLTFQSDSTVSQSGSHALGDRIYSFSSTSSSES LN 85	
Dd	107	TTTASTPEFTISLTSLKTSTDSTSEPRISTTDTKDTTPEDVPSSSTDQSSTSPHETTR 166	
Qy	86	SSAPRGERSTAGIS-YGOVRGTATIEORTS-----SDHTDHLYLSSTFTFKGERALLST 137	
Dd	167	DTYTEGTSEDSTSYGSTERRSSPKPTSEFSTESTESDTESTRSSETIS-----SIE 219	
Qy	138	DNSSSDIVESSTSYIKINSHSSEYSSFSHA-----QTERNISSYDGGEYAQPSTESP 192	
Dd	220	TNSSTSPV--STRPEVDSTSSGSNETTESDGTMTTVFTTTKDDTSIVSGDSNGSGSTSEF 277	
Qy	193	LHNTSNLPSPYPTTMPNTSVVLTDADFVSDSSSSSSSSSSSS-SSSGPPLPSPVSQSHH 251	
Dd	278	KNTETTTPGSGTVSEPSSSRSDLDSSSVSDRSTDQRRTTEIGLQGPI L-----327	
Qy	252	LFSSILPSTRASVHLKLKSTDASTPWSSPSPLPVSLTWTSTAPLSVSQTTL PQSSSTPV 311	
Dd	328	-----SDSSNNPDPTTSALTSCGTTSTGRASSADDP TTTGCST--367	
Qy	312	LPRARETPVTSFOSTMTWS---FMTMLHSSQATDLKSQSFPHOEKVITESKP SLVS-LP 367	
Dd	368	-----TSGTASTTSGSLFSTSLGSSQSPGSSVSFTP-----GPSITGIS 408	
Qy	368	TESTKAVTTHNPSPPLPSTESSTEQTLPATSTNLQAQMSPTTTTTILKTSOPLM TTFC-TLS 426	

Db	409	QSTTSGPPTTS	--EPSTTSGTSDTSGPSTT	-SGPSTTLLGTGTQSTTSGP	-STTPGSTIS	464		
Qy	427	ST	--ASLVTGP	-----IAVOTTAGKQLSLTHPEILVQI	ISTEGGISTERNRVIVDATTGL	479		
Db	465	TTASATTS	SGPSTSSGTVSTTSGQSTSSGTTKSTTSGP	TSSGPTSSGPSTVSERT	-----	516		
Qy	480	IPLTSVPTSAKEMTK	LGVTAEYSPASRSIGTSPSPOTTVVSTAEDLAPKSA	TAFVOSST	539			
Db	517	LSTTSGPSTTSGPSTTSGTVSTTPGAST	--TSGTSTQSTTSGPSTSSGPSTASRSTVST	574				
Qy	540	QSPITLS	-----SSASVNSCAVNPCLHNGECVADN	TSRGYHCRCPSPHQGD	CSV	589		
Db	575	SGPSTTSGPSTTSGPSTTSGTSTTSGPSTTSGKNISTVSGK	-----LTG	STTSA	625			
Qy	590	DVNECLSNPCPSTATC	NNTOGCSFFICKCPVGQLEKGCICNLVR	TFTVEFKLKRFL	NTTIVE	649		
Db	626	TISAFGNGVFTSKP	NSSGGTTSSGKNFSQNTTSAANGTTQAVNNGKSGSTL	PTN	SSS	685		
Qy	650	KHSDQEVENEITK	LNMCFSA	LPYSIRSVTHASRESNA	VISLQTFSLAS	NWT-----704		
Db	686	GSDSSSTSPSTFTVSYNNTNSSI	-TCIEPSLNLT	TYTSSPTSSQIKSSYSV	GELITQVCPQ	744		
Qy	705	--LFDLA	--DRMOKCVNSCK	--SSA	EV	726		
Db	745	SYVEIALQPVKIYK	CKLTKWAGSPENC	773				
RESULT	13							
Q9H195	ID	Q9H195	PRELIMINARY;	PRT;	901 AA.			
AC	Q9H195;							
DT	01-MAR-2001 (TremBLrel. 16, Created)							
DT	01-MAR-2001 (TremBLrel. 16, Last sequence update)							
DT	01-DEC-2001 (TremBLrel. 19, Last annotation update)							
DE	MUC3B MUCIN (FRAGMENT).							
GN	MUC3B.							
OS	Homo sapiens (Human).							
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
OX	NCBI_TaxID=9606;							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RA	Pratt W.S., Crawley S., Hicks J., Ho J., Nash M., Kim Y.S., Gum J.R.,							
RA	Swallow D.M.;							
RT	"Multiple transcripts of MUC3: Evidence for two genes, MUC3A and							
RT	MUC3B.";							
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.							
DR	EMBL; AJ291390; CAC19572.1; -.							
DR	InterPro; IPR000561; EGF-like.							
DR	InterPro; IPR000082; SEA.							
DR	SMART; SM00181; EGF: 2.							
DR	SMART; SM00200; SEA: 1.							
DR	PROSITE; PS00022; EGF_1; UNKNOWN_1.							
DR	PROSITE; PS50024; SEA; 1.							
FT	NON_TER	1						
SQ	SEQUENCE	901 AA;	96192 MW;	F98FC817494ECD99	CRC64;			
Query Match		6.9%;	Score 337.5;	DB 4;	Length 901;			
Best Local Similarity		22.0%;	Pred. No. 1.8e-12;					
Matches	211;	Conservative	117;	Mismatches	289;	Indels 341;		
Gaps								
Qy	132	ALLSITDSSSDIVESSTSYKISNSHSEYSFSFHAOTERNIS	SYDCEY	AQPSTESP	191			
Db	6	SIITETTSHTSFSSSTHSTVSSSTTATPGTTAE	GTSTPS	-----SPSS	LST	59		
Qy	192	VLHTSNLPSVTP	TINMPTSVL	TDIAEFYSDSSSSSSSSSSSSSGG	PLPLP	SVSQSHH	251	
Db	60	DIPTSLRLT	PLSLSTSLTTTDLPSIPTDISS	LP	PIHISSPS	IQSTETSS	LVG	119
Qy	252	LFSSILPSTRASV	HLKSTSDAS	-----TPWSSSP	-----	SPLP	VSLT	289
Db	120	TTPSTMGT	VARLSTRENTPISSFSTSVVPTPTT	QAPVLM	SATG	TQTS	SPVPTVTF	179

Query Match	6.9%	Score 337.5;	DB 4;	Length 901;
Best Local Similarity	22.0%	Pred. No. 1.8e-12;		

Qy	132	ALLSITDONSDDIVESSTSVIKISNSSHSEYSFSHAOTERNISGYDGEYAOQPSTSP	191
Db	6	SIITETTSHTSFSSSTHVSSSSTTATSPFTTAETGVSTPS-----SPSLSLT	59
Qy	192	VLHTSNLPSYPTTNMPTSVGLTDAEFVSDSSSSSSSSSSSSSSSSSSSSSSSSSSSS	251
Db	60	DIPTSLRTLTPLSLSTSLTTTDLPSITDSSLPTPIHIISSSPSQSTETSSLVG	119
Qy	252	LFSSILPSTRASVHLLKSTSDAS-----TPNSSSP-----SPLPVSLT-	289
Db	120	TTSPMTGVRATLRSTNTPISSFSSTSVVTPETPTTOAPPVLMASATGOTSPVPTVTF	179

```
QY 290 -----TSTAPLSVSQTLTPOSSS-----TPVLPRAREPTVTSF----- 323
Db 180 GMSDSSSTLHLLTPSALKIMSTOFFPISTHSSTLQTPSIP-SLQSLSTSEFTT 238
QY 324 -----QTST---MTSFMTLHSSQTADLKSQS-----TPHQEKVI-----TE 357
Db 239 ESFTRGSTNAITLSTFIWSSTPIIMSSSPSSASITPVFATTIHSVPSPPIFSTE 298
QY 358 S-KPSLSVLPTSTKATVNTSPLPPLSTESSSTEQTLPATSTNLAQSP---TFTTIL- 412
Db 299 NVGASITAFPSLSSTSTSPSSSLTALTET-PPSYISLPSPTPCPGTITIVP 357
QY 413 -----KTSQPLMTTPGL-SSTASLVTPG--IAVQTAGKQL-----SLTHP 451
Db 358 ASPTDPCVEMDPSEATSP--TTPLVFFPTEMTVCPSISNQTLATHMTSSMT-P 414
QY 452 E---ILVLPQISTEGGIST-----ERNRVIDATGTL--IPLTSVPT 487
Db 415 ESESIIPNASSSTGTCTVPTNTVFTSTRLPTSETWLSNNSVIPPTPLPGVSTIPLTKPS 474
QY 488 SAKEMTKLGVTAIYSPASSLGTSPTQTVVSTAEDLAPKSAFVQSTQSTPTLSS 547
Db 475 SS---LPTIL-----RTSSKTHPSPPTARTSQT-----VATQTPPTLT- 512
QY 548 SASYNCAVNPCLH-----NGECVADNTSRGYHCRCPSPSQGDCCSDVDNECLS 596
Db 513 -----TCRTPTITSMMTQSTLTATTACTDNGGTWEOGQACLPFGSGDCOLQTR---- 563
QY 597 NPCSTATCANN--TOGSFICKCPGVQLEKIGICNLVFTVTEFKLKTFLNTTVE---KH 651
Db 564 -----CONGGOWDLKQCP-----STFYGSCEFAVEQ 592
QY 652 SDLOVEVNETKTNLMCFALSPIRSTVHASRESNAV-----V 690
Db 593 VLDLVTEVGMESVVDQFQSPDLNDNTSQAIRDFNTFNQMKIFADWQGFKGVEI 652
QY 691 ISLQTTESLASNTFLD-----LADMOKC-----VNSKSSAEVCCOLLG 730
Db 653 LSLRANGSIVVDYLVLEMPFSPQLESEYEQVKTTLKEGLQNASODANSQDQALC----- 708
QY 731 SQRIFRAGSLCKRKSPECCKDSICTDLQVALCO--CKSGVQF----- 774
Db 709 -----FKPDSIKVNNNSK-----TELTPAICRAAPTGYEEFYPLVATRLRCVT 755
QY 775 -----NKMDHSCRACEDGYRLENETCMSCPFGLGLNC-----GNPYQLITVV 817
Db 756 KCTSGLDNAID-----CHQG-----OCVLETSGPACRCYSTDTHWFGSPCEVTVH 801
QY 818 IAAAGGGLLLILGIALIVTCRKNKNDISKLIFKSGDFQMSPYAEYKPNRPSQEWGRE 875
Db 802 WRALVGG--LTAGAALLVL-----LLALG-----VRAVRSEWNGRQ 836

RESULT 14
Q939N5 PRELIMINARY; PRT; 3072 AA.
ID AC Q939N5
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE PLATELET BINDING PROTEIN GSPB.
GN GSPB.
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1302;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M99;
RA Bensing B.A., Sullam P.M.;
RT "An accessory sec locus of Streptococcus gordonii is required for
RT export of GspB and for platelet binding.";
```

```
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY028381; AAL13053.1; -
SQ SEQUENCE 3072 AA; 285770 MW; 0BJ48372697CF7F2 CRC64;
```

```
Query Match 6.9%; Score 336; DB 2; Length 3072;
Best Local Similarity 24.7%; Pred. No. 1.1e-11;
Matches 140; Conservative 112; Mismatches 281; Indels 34; Gaps 10;
```

```
QY 1 MSQTEVSRSVAPMRGGEITAHWLLTNTSTTSADVTS-----SASYEGVNASVLVT 51
Db 637 ISSASVSASTSISTSGSVAS---ESASTSSVSASESASTSASVSASESASTSASV 693
QY 52 QFSQSTVQSGSHPALGDRSYSESSSTSSSLNSLAPRGRSITAGISYGVGRGTAEQR 111
Db 694 STSASTSASVSASTS-ASTSASTSASTSASTSASTSASTSASTSASTSASTSASV 752
QY 112 TSSDHTHTYLSSTFTTKGERALLSITDNSSSDIVESTSYIKINSNHSYSFSF---H 168
Db 753 ASTSASTSASVSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASV 812
QY 169 AQTERSNISSYDGEYAQPSTESPVLHTSNLPSYPTINMPNTSVVLDTDAEFVSDSSSS 228
Db 813 ASTSASTSASVSAS-ASASTSASVSASTS-----ASTSASVSASAS--ASTSASV 859
QY 229 SSSSSSSSGPPLPLPSVSQSHHLFSILPSTRASVHLLKSTSDASTPWSSSPPLPVSL 288
Db 860 SASVSASTSASVSASESASTSASVSASESASTSASVSASESASTSASVSASESASTS 919
QY 289 TTSAPLSVSQTLTPOSSSTPVLPRAETPVTSFQSTMTSFTMTLHSSQTADLKOST 348
Db 920 SASTSASTSAS-VSASESASTSASVSASESASTSASVSASESASTSASVSASESAST 975
QY 349 PHQEKVITESKPSLIVSLPTSTKAVTNSPLPPSLTESSEQTLPATSTNLAQSPFTT 408
Db 976 SASVSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSASTSAST 1035
QY 409 TTILKTSQPLMTPTGLSTLSASLTGPIAVQTAGKQLSLTHPEILVPQISTEGGISTER 468
Db 1036 SASVSASESASTSASVSASTSASTSASTSASTSASTSASTSASTSASTSASTSAS-ESASESAS 1094
QY 469 NRVIYDATTGLPILTSVPTSAKEMTKLGVTAIYSPASRSLGTSPTQTVVSTAEDLAP 528
Db 1095 TSASVSASESASTSASVSASESSTSSASTSASVSASESSTSSASTSASVSASESASTSASVSASESASTS 1154
QY 529 KSATFVQSSSQSPSTLSSASVNSCA 555
Db 1155 TSASESASESASTSASVSASESASTSA 1181

RESULT 15
Q9KWR3 PRELIMINARY; PRT; 2178 AA.
ID AC Q9KWR3
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE STREPTOCOCCAL HEMAGGLUTININ.
GN HSA.
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1302;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DLL;
RA Takahashi Y., Konishi K., Yoshikawa M.;
RT "Cloning and characterization of the gene encoding a hemagglutinin of
RT Streptococcus gordonii DLI.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB029393; BAA97453.1; -
DR InterPro; IPR004089; Chemotaxis_transducer.
DR InterPro; IPR001899; Gram_pos_anchor.
```

DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SQ SEQUENCE 2178 AA; 203507 MW; 75CC27E27F41DA8C CRC64;

Query Match 6.9%; Score 334.5; DB 2; Length 2178;
Best Local Similarity 22.0%; Pred. No. 8.5e-12;
Matches 177; Conservative 145; Mismatches 367; Indels 115; Gaps 19;

QY 2 SOTETVSRVAPMRGGTEITAHLLTNS-TTSDADVTGS-----SASYPEGYNASVLT 51
Db 1434 SESAYTSASVASESGSTASVSASESTASVSASESTASVSASESTASVSASESTASVSACA 1493
QY 52 QFSDST---VQSGSHALGDRSYSESSSTSSSESLNS-----SA 88
Db 1494 SASSTASVSASGASTASVSASESTASVSASTASVSASESTASVSASESTASVSASA 1553
QY 89 PRGERSIAGISYQGVRCGTAEQRTSSDHTHTYLSSTFTTGERALLSITDNSSSDIV-- 146
Db 1554 SESASTASVSASESTASVSASTASVSASTASVSASESTASVSASESTASVSASA 1613
QY 147 -ESSTSVIKISNSSHSSEYSEFSHAQTERSNISSYDGEYAQPSTESPVLHTSNLPSYPTI 205
Db 1614 SESSTASVS-ASESTASVSASESTASVSASESTASVSASESTASVSASESTASVSVS 1672
QY 206 NMPNTSVYLDTADEFVSDSSSSSSSSSSSGPPPLPSPVSOQHHLFSSILPSTRASVH 265
Db 1673 ASESASTASVSASESASSASVSASKASMSASVLAASESTASVSASESTASVSASVS 1732
QY 266 LLKSTDASTPMSSPPLPVSLTTSAPLSVSQTTLPOSSSTPVLPRARETPVTSFQT 325
Db 1733 ASESASTASVSASESTASVSASESTASVS-VSASESTASVSASESTASVSASTASV 1791
QY 326 STWTSFMTLHSSQTADLKQSOTPHQKVITESKPSLVSILPTSTKAVTNTSNLPPSLT 385
Db 1792 SASTS-----ASTSASVSASESTASVSASESTASVSASESTASVSASESTASVSAS 1842
QY 386 ESSTEQTLPATSTNLAQMSPTFTTILKIQPLMTTPGTLSTASLVGTPIAVOTTAGKQ 445
Db 1843 ESVSASESTASVSASTASVSASTASVSSESESTASVSASESTASVSASESTASVS 1902
QY 446 LSLTHPE--ILVPOISTEGGISTERNRVIDATWGTLPLTSVPTSAKEMTKLGVTAEYS 503
Db 1903 ASVSASESTASISASESSSTAS---VSASESTASVSASTSTSTSTASVSASESA 1959
QY 504 PASRSLGTSPPQTVVYTAEDLAPKATFAVQSTQSTPTLSSASVNSCAVNPCLHNG 563
Db 1960 STSASVFASESTASVSASESTASVSASTASVSASTASVSASESTASVSASESTAS 2011
QY 564 ECVADNTRSGYHCRPCPPSQWQDDCSVDVNECLSNPCP-STATCNNTQGSFICKCPVGYL 622
Db 2012 -----SISASESTASISASESSSTASVSASTASVSASTSA 2047
QY 623 EKGICNLVRFVTEFKLKRFTLNTTVEKHSDLQEVENEITKTLNMCFSALPSYIRSTVHA 682
Db 2048 SVSASESTSTSVISASESVISSTSVSQMSVSE-----SLSLSVSTSTLHSQLNG 2098
QY 683 SRESNAVVISLOTFTSLASNV--TLFD-----LADRMQK-----CVNSCKSSAEVCQ 727
Db 2099 IYSELNSLSLSELSMSQSLQSLSQSTSATQSMHDIRISKQLPRTGESEKASILA 2158
QY 728 L-LGSQRRIFRAGSLCKRKSPCED 750
Db 2159 LGTGLGLAFK-----KRKNESE 2177

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 20:05:41 ; Search time 955.43 Seconds
(without alignments)
12492.793 Million cell updates/sec

Title: US-09-840-746-2

Perfect score: 6952

Sequence: 1 gttcgatgaagaattgccg.....tactgggaagctatagggtg 6952

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /SID55/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
- 2: /SID55/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
- 3: /SID55/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
- 4: /SID55/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
- 5: /SID55/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
- 6: /SID55/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
- 7: /SID55/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
- 8: /SID55/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
- 9: /SID55/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
- 10: /SID55/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
- 11: /SID55/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
- 12: /SID55/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
- 13: /SID55/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
- 14: /SID55/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
- 15: /SID55/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
- 16: /SID55/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
- 17: /SID55/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
- 18: /SID55/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
- 19: /SID55/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
- 20: /SID55/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
- 21: /SID55/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
- 22: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
- 23: /SID55/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
- 24: /SID55/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5098.4	73.3	6407	20	AXA26246
2	5098.4	73.3	6407	20	AAV81825
3	5098.4	73.3	6407	21	AAA88578
4	5098.4	73.3	6407	21	AAZ89796
5	5098.4	73.3	6407	21	AAZ88007
6	5042.4	72.5	6407	17	AAZ36035
7	3952.8	56.9	3975	22	AAK94374
8	3092.8	44.5	3116	22	AAK94374
9	3080	44.3	3170	21	AAAF18333

10	831	12.0	3075	22	AAAL26437	Human breast cancer
11	708.2	10.2	751	22	AAK92040	Human CDNA 5'-end
12	708.2	10.2	751	22	AAK93703	Human CDNA clone r
13	703	10.1	869	21	AAH31088	Human colon cancer
14	500.8	7.2	537	22	AAAL34960	Human musculoskele
15	480	6.9	546	22	AAK92794	Human CDNA 3'-end
16	471	6.8	616	22	ABA06457	Human CDNA SEQ ID
17	471	6.8	616	22	AAK28946	CDNA encoding for
18	471	6.8	616	22	AAK26918	Human CDNA encoding
19	471	6.8	616	22	AAK31583	CDNA encoding nove
20	470.4	6.8	584	21	AAK79999	Human colon cancer
21	386.4	5.6	400	21	AAH30808	Human colon cancer
22	385.8	5.5	400	22	AAK37279	Novel human diagno
23	357	5.1	391	21	AAH30820	Human colon cancer
24	351.4	5.1	781	22	AAK26846	Human CDNA encoding
25	331.2	4.8	351	21	AAA50236	DNA encoding autoa
26	247.2	3.6	252	22	AAK25327	Human breast cancer
27	231	3.3	248	20	AAV88062	EST clone FS173.
28	230.4	3.3	403	22	AAK07581	Human breast cancer
29	191.8	2.8	195	22	AAAL16484	Human breast cancer
30	152	2.2	349	22	AAAL19125	Human breast cancer
31	150	2.2	343	22	AAK25960	Human breast cancer
32	136	2.0	333	22	AAK08215	Human breast cancer
33	82.6	1.2	344	22	AAK56290	Human immune/haema
34	57.4	0.8	651	23	AAK84956	DNA encoding novel
35	56.8	0.8	824	23	AAK66529	DNA encoding novel
36	56.4	0.8	1072	23	AAK90738	DNA encoding novel
37	56.2	0.8	1050	22	AAK21685	Mutational hot spo
38	56.2	0.8	2849	22	AAK21684	Human retinitis pi
39	56.2	0.8	3489	21	AAA30290	Kaposi's sarcoma-a
40	56.2	0.8	3489	22	AAK82901	Nucleotide sequenc
41	56.2	0.8	32207	20	AAV73805	KSHV LUR DNA (nucl
42	56.2	0.8	137507	19	AAV19941	KSHV long unique c
43	56	0.8	583	22	ABA50512	Human breast cell
44	56	0.8	583	22	ABA68463	Human foetal liver
45	56	0.8	583	22	ABA35452	Probe #13918 for g

ALIGNMENTS

RESULT 1

AXA26246
ID AXA26246 standard; CDNA; 6407 BP.
XX
AC AXA26246;

XX
XX 24-MAY-1999 (first entry)

XX
DE CDNA sequence of rchd528 gene.

XX
KW Fingerprinting gene; rchd502; transmembrane protein; cardiovascular;
KW fingerprint/target gene; up-regulated; endothelial cell; shear-stress;
KW atherosclerosis; ischemia; reperfusion; hypertension; restenosis; ds.

XX
OS Homo sapiens.

XX
PN US5882925-A.

XX
PD 16-MAR-1999.

XX
PF 09-FEB-1996; 96US-0599654.

XX
PR 09-FEB-1996; 96US-0599654.

XX
PR 10-FEB-1995; 95US-0386844.

XX
PR 07-JUN-1995; 95US-0485573.

XX
(MILL-) MILLENNIUM PHARM INC.

XX
PI Falb DA;

XX
XX WPI; 1999-214071/18.

XX
XX P-PSDB; AAW95160.

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX
XX

XX New polynucleotides consisting of residues 1-1929 of the rchd502
PT gene - are differentially expressed in cardiovascular disease
PT states, and can therefore be used to treat and diagnose
PT cardiovascular diseases
XX
XX Disclosure; Fig 30A; 121pp; English.
XX
CC The invention relates to a rchd502 target/fingerprint gene encoding a
CC transmembrane protein. The invention provides cDNAs contained in plasmids
CC pFCHD502SF (ATCC 69981) and pFCHD502SJ (ATCC 69982) that encode the
CC rchd502 polypeptide, and are differentially expressed in cardiovascular
CC disease states. Cultured genetically engineered host cell containing the
CC rchd502 polynucleotides in operative association with a nucleotide
CC regulatory element are used for producing a polypeptide rchd502 gene
CC product. Identifying that the fingerprint/target gene rchd502 is
CC differentially expressed (up-regulated) by endothelial cells subjected
CC to shear-stress, provides a tool for the diagnosis and treatment of
CC cardiovascular disease e.g. atherosclerosis, ischemia/reperfusion,
CC hypertension, restenosis. The fingerprint gene is useful for testing the
CC efficacy of candidate drugs in basic research and in clinical trials and
CC or imaging of a diseased cardiovascular tissue. The gene may also be
CC used in screening for ligands of target gene product receptor domains, as
CC well as antagonists of the ligand-receptor interaction.
XX
SQ Sequence 6407 BP; 1594 A; 1762 C; 1520 G; 1509 T; 22 other:

Query Match 73.3%; Score 5098.4; DB 20; Length 6407;
Best Local Similarity 94.3%; Pred. No.0;
Matches 5412; Conservative 21; Mismatches 8; Indels 301; Gaps 2

QY 1 gttcgtgaaagaattgcgcgttttcaaacaaagagtgaacagcctcgagatgggaac 60
DB 666 gttcgtgaaagaattgcgcgttttcaaacaaagagtgaacagcctcgagatgggaac 725

QY 61 agagagggcgatggggctgtcagaagaatggactgtgcacagccaaaggccaccactc 120
DB 726 agagagggcgatggggctgtcagaagaatggactgtgcacagccaaaggccaccactc 785

QY 121 ggcttagagcccttccttcttcctgcttggagatggagagctgaccacgctcttag 180
DB 786 ggcttagagcccttccttcttcctgcttggagatggagagctgaccacgctcttag 845

QY 181 gaagagaatctctcaggaccagatctcctcgtgctgattctacaggacagcagcttc 240
DB 846 gaagagaatctctcaggaccagatctcctcgtgctgattctacaggacagcagcttc 905

QY 241 ctctcctctcttagaccttctcactctccttgaagatcacagaagcttaacactccac 300
DB 906 ctctcctctcttagaccttctcactctcgtgaaagtacagaagcttaacactccac 965

QY 301 tggcctccagagctcctcagtcagtcacaacaaagacaatgcatttgcctaccgtgtccac 360
DB 966 tggcctccagagctcctcagtcagtcacaacaaagacaatgcatttgcctaccgtgtccac 1025

QY 361 tgatggtgcccagagaacgctcgatctttgacggtcagctggagacctgtgagcaagac 420
DB 1026 tgatggtgcccagagaacgctcgatctttgacggtcagctggagacctgtgagcaagac 1085

QY 421 agaaggtctcccaaggactccagaattgccacgacttcactcagctctcttcttccac 480
DB 1086 agaaggtctcccaaggactccagaattgccacgacttcactcagctctcttcttccac 1145

QY 481 ctctcagctgggaatcgagaagaaacagtagagtaactgggaatccaggggatgaggaatt 540
DB 1146 ctctcagctgggaatcgagaagaaacagtagagtaactgggaatccaggggatgaggaatt 1205

QY 541 cattgaaccatccacagaaaaatgaatttggactacgctcttgcg-tggcacaatgattc 599
DB 1206 cattgaaccatccacagaaaaatgaatttggactacgctcttgcg-tggcacaatgattc 1265

QY 600 cccaaacctttgagaaacatcacgcttgcacagcagctctgaggtgcacaaatggaatgccat 659

Db 2346 gcectgcccctgtgtcacaatcccaacatttattttcattcatttccatcaaccag 2405
 QY 1440 ggcectgtgtcatctactaaagttaacctctgtatgtaoatccacaccatggtctctccacc 1499
 Db 2406 ggcectgtgtcatctactaaagttaacctctgtatgtaoatccacaccatggtctctccacc 2465
 QY 1500 atcacctttaccagttatcttaagacatctacatctgcccactttctgtctcacaac 1559
 Db 2466 atcacctttaccagttatcttaagacatctacatctgcccactttctgtctcacaac 2525
 QY 1560 aaccttggccagttcatcttctaccctctccctggccagggcgaagggagactcctgtgac 1619
 Db 2526 aaccttggccagttcatcttctaccctctccctggccagggcgaagggagactcctgtgac 2585
 QY 1620 ttcatttcagacatcaacaatgacataattcatgacaatgtctccatagtagtcaaaactgc 1679
 Db 2586 ttcatttcagacatcaacaatgacataattcatgacaatgtctccatagtagtcaaaactgc 2645
 QY 1680 agaccttaagagccagagcaccacacacaagagaaagtctattacagaatcaaaagtcacc 1739
 Db 2646 agaccttaagagccagagcaccacacacaagagaaagtctattacagaatcaaaagtcacc 2705
 QY 1740 aagcctgggtgtctctgcccacagagtcacccaaagtgttaacacaaaactctctcttggc 1799
 Db 2706 aagcctgggtgtctctgcccacagagtcacccaaagtgttaacacaaaactctctcttggc 2765
 QY 1800 tccatcttaacagagtcctccacagagcaaaacccctccacagccacaaagcaacaaacttagc 1859
 Db 2766 tccatcttaacagagtcctccacagagcaaaacccctccacagccacaaagcaacaaacttagc 2825
 QY 1860 acaaaatgtctccaaacttcaaaactacattctgaagacactctcagcctcttatgaccac 1919
 Db 2826 acaaaatgtctccaaacttcaaaactacattctgaagacactctcagcctcttatgaccac 2885
 QY 1920 tccctggccacctgtcaagcacagcatctctgttgcactggccctatagccgtacagactac 1979
 Db 2886 tccctggccacctgtcaagcacagcatctctgttgcactggccctatagccgtacagactac 2945
 QY 1980 agctgaaacacagctctcgtgaccatcctgaaatactagtctccaaatctcaacaga 2039
 Db 2946 agctgaaacacagctctcgtgaccatcctgaaatactagtctccaaatctcaacaga 3005
 QY 2040 agdggcactagcacagaaaggaacaggatgattgttggaagtctaccactggattgatccc 2099
 Db 3006 agdggcactagcacagaaaggaacaggatgattgttggaagtctaccactggattgatccc 3065
 QY 2100 tttagccagtgtagccacatcagcaaaaagaaatgaccacaaaagtctggcggttacacagaga 2159
 Db 3066 tttagccagtgtagccacatcagcaaaaagaaatgaccacaaaagtctggcggttacacagaga 3125
 QY 2160 gtacagccagcttcacgttccctcggaaacatctccttctcccaaacacacagttgttcc 2219
 Db 3126 gtacagccagcttcacgttccctcggaaacatctccttctcccaaacacacagttgttcc 3185
 QY 2220 cacggctgaagacttggtctcccaaatctgcacacttctggttcagagcagcacacagtc 2279
 Db 3186 cacggctgaagacttggtctcccaaatctgcacacttctggttcagagcagcacacagtc 3245
 QY 2280 accaaacaaactgtctcttctcagcctcagtcacacagctgtgtgtgtgaaccccttcttcca 2339
 Db 3246 accaaacaaactgtctcttctcagcctcagtcacacagctgtgtgtgtgaaccccttcttcca 3305
 QY 2340 caatggcggaatggctgcagagaaaacccagccgtggtaccactcaggtgccgccttc 2399
 Db 3306 caatggcggaatggctgcagagaaaacccagccgtggtaccactcaggtgccgccttc 3365
 QY 2400 ctggcaggggagtgattgcaagtgtgattggaatgagtgctgtcgaacccctgcccac 2459
 Db 3366 ctggcaggggagtgattgcaagtgtgattggaatgagtgctgtcgaacccctgcccac 3425
 QY 2460 cacagccagtgcaacaataactcagggatccttcttatctgcaaatgcccggttgggtacca 2519
 Db 3426 cacagccagtgcaacaataactcagggatccttcttatctgcaaatgcccggttgggtacca 3485

QY 2520 gttggaaaaaggagatgcaatttggtagaacctctgtgacagagtttaaatataagag 2579
 Db 3486 gttggaaaaaggagatgcaatttggtagaacctctgtgacagagtttaaatataagag 3545
 QY 2580 aacttttcttaatacacaactgtgaaaaaacattcagacactcaagaagttgaaaatgagat 2639
 Db 3546 aacttttcttaatacacaactgtgaaaaaacattcagacactcaagaagttgaaaatgagat 3605
 QY 2640 caccaaaacgttaaatatgtgtttttcagcgttaccttagttacatccgatctacagtcca 2699
 Db 3606 caccaaaacgttaaatatgtgtttttcagcgttaccttagttacatccgatctacagtcca 3665
 QY 2700 cgctctaggaggtccaacgcggtgtgtatctcactgcacaaacacttttccctggcctc 2759
 Db 3666 cgctctaggaggtccaacgcggtgtgtatctcactgcacaaacacttttccctggcctc 3725
 QY 2760 caatgtgacgtctattgacctgctgatatgtagtcagaaatgtgtcaactccttgcgaagtc 2819
 Db 3726 caatgtgacgtctattgacctgctgatatgtagtcagaaatgtgtcaactccttgcgaagtc 3785
 QY 2820 ctctgctgaggtctgccagctcttgggatctcagagcggtatcttttagagcgggcagctt 2879
 Db 3786 ctctgctgaggtctgccagctcttgggatctcagagcggtatcttttagagcgggcagctt 3845
 QY 2880 gtgcaagcggaagagtcctccgaatgtgacaaagacacacotccatctgcaactggagcgg 2939
 Db 3846 gtgcaagcggaagagtcctccgaatgtgacaaagacacacotccatctgcaactggagcgg 3905
 QY 2940 cgttgcctgtgcccagtgcaagtcgggatacttccagtccaacaagatggaccactcctg 2999
 Db 3906 cgttgcctgtgcccagtgcaagtcgggatacttccagtccaacaagatggaccactcctg 3965
 QY 3000 ccagagcatgtgaagatgggatataggtctgaaaaatgaaacctgcaatgagttgcccattgg 3059
 Db 3966 ccagagcatgtgaagatgggatataggtctgaaaaatgaaacctgcaatgagttgcccattgg 4025
 QY 3060 ccttgggtgtcacaactgtgaaacccctatcagcttatcactgtgtgtgacgagccgc 3119
 Db 4026 ccttgggtgtcacaactgtgaaacccctatcagcttatcactgtgtgtgacgagccgc 4085
 QY 3120 gggaggtgggctcctgctcatctaggtcgcactgattgttacctgttgcgaaagaa 3179
 Db 4086 gggaggtgggctcctgctcatctaggtcgcactgattgttacctgttgcgaaagaa 4145
 QY 3180 taaaaatgacataagaacaaactcatcttcaaaagtggagatttccaaatgtcccatagc 3239
 Db 4146 taaaaatgacataagaacaaactcatcttcaaaagtggagatttccaaatgtcccatagc 4205
 QY 3240 tgaatcccccaaaaatctcctgcgcacaaagattggggccgagaagctattgaaatgcataga 3299
 Db 4206 tgaatcccccaaaaatctcctgcgcacaaagattggggccgagaagctattgaaatgcataga 4265
 QY 3300 gaatggaagtaccaaaaaactcctccagatgacggatgtgttactactcgcctacaaagt 3359
 Db 4266 gaatggaagtaccaaaaaactcctccagatgacggatgtgttactactcgcctacaaagt 4325
 QY 3360 aaggaatccgaacttgaaacgaacggactctacccggcctacactggaactgcccagatc 3419
 Db 4326 aaggaatccgaacttgaaacgaacggactctacccggcctacactggaactgcccagatc 4385
 QY 3420 acggcattcttgattttcccggacagataaacccgtcttttcatcagtagtgaagacag 3479
 Db 4386 acggcattcttgattttcccggacagataaacccgtcttttcatcagtagtgaagacag 4445
 QY 3480 aagaagagactacttttaagtccagagagagagggactcatctgctgtgagcagtcacc 3539
 Db 4446 aagaagagactacttttaagtccagagagagagggactcatctgctgtgagcagtcacc 4505
 QY 3540 tgggacctctgctcagaggaccacacagaggtcgcgccaggatttgcgggagccac 3599
 Db 4506 tgggacctctgctcagaggaccacacagaggtcgcgccaggatttgcgggagccac 4565

QY 3600 gctgagtcgcaagcaggaaagggacaggcatgcggggcgctgaccacagtcggaggagaca 3659
 Db 4566 gctgagtcgcaagcaggaaagggacaggcatgcggggcgctgaccacagtcggaggagaca 4625
 QY 3660 ggtggatgtggaaccacagcgtcctcattcagcacctttgttttactgtgaacgtgaat 3719
 Db 4626 ggtggatgtggaaccacagcgtcctcattcagcacctttgttttactgtgaacgtgaat 4685
 QY 3720 gtcggccagtcataagagagtcctctgagtcgactgcaccatggcactggccacaggcg 3779
 Db 4686 gtcggccagtcataagagagtcctctgagtcgactgcaccatggcactggccacaggcg 4745
 QY 3780 actattagccaggccagaccactagacttcagtcgagggaacctggtttcccttcgtttg 3839
 Db 4746 actattagccaggccagaccactagacttcagtcgagggaacctggtttcccttcgtttg 4805
 QY 3840 caotttagtaaaattgggtgggaagtttcotttttggatcgtttttgagactgtttccagaaa 3899
 Db 4806 caotttagtaaaattgggtgggaagtttcotttttggatcgtttttgagactgtttccagaaa 4865
 QY 3900 gaaggcttcctttcccgagacacttccatagcgagcaaatctggtgatttcatttgcagcaa 3959
 Db 4866 gaaggcttcctttcccgagacacttccatagcgagcaaatctggtgatttcatttgcagcaa 4925
 QY 3960 aatactggctgttaattattttctgcgcagcgctgcgtgctaaacaacagatgagga 4019
 Db 4926 aatactggctgttaattattttctgcgcagcgctgcgtgctaaacaacagatgagga 4985
 QY 4020 tggagctaacactgaagtcgaagatgctgccattgaacggacagtgcttttcatagtttt 4079
 Db 4986 tggagctaacactgaagtcgaagatgctgccattgaacggacagtgcttttcatagtttt 5045
 QY 4080 ctagggtgtcttatgctacagtttccaaagccagcccccacagtcgagaaatgctgaaggc 4139
 Db 5046 ctagggtgtcttatgctacagtttccaaagccagcccccacagtcgagaaatgctgaaggc 5105
 QY 4140 acgcacacaaactgcgaatgtgttttttaagtcgaagtgacacatgtatttaagatattttt 4199
 Db 5106 acgcacacaaactgcgaatgtgttttttaagtcgaagtgacacatgtatttaagatattttt 5165
 QY 4200 tttaaaatcctcttgcaagttaaaatcacttttttcaaaacaaagcctggatcagggcnaaac 4259
 Db 5166 tttaaaatcctcttgcaagttaaaatcacttttytcaaaacaaagcctggatcagggcnaaac 5225
 QY 4260 aacttatatttggcttagctgagcctcagcagcgagattgcagcgaggggggcactttt 4319
 Db 5226 aacttatatttggcttagctgagcctcagcagcgagattgcagcgaggggggcactttt 5285
 QY 4320 tcattcaatgagggccacagcctggggcctgggaactctgataccatttgtggaggccagagg 4379
 Db 5286 tcattcaatgagggccacagcctggggcctgggaactctgataccatttgtggaggccagagg 5345
 QY 4380 cagctcgctatggagagaaatgcaaaactgaacgaggtttccaccactaggaagca 4439
 Db 5346 camctcgctatggagagaaatgcaaaactgaacgaggtttccaccactaggaagca 5405
 QY 4440 gcttgttgagccctcgacgtgtagtggtttagaggatgggctgaaatgagcagggttaga 4499
 Db 5406 gcttgttgagccctcgacgtgtagtggtttagaggatgggctgaaatgagcagggttaga 5465
 QY 4500 ttctctgatacaacagtcgttttgggaagtcgtgtgattcctgaggaagaacaggagacc 4559
 Db 5466 ttctctgatacaacagtcgttttgggaagtcgtgtgattcctgaggaagaacaggagacc 5525
 QY 4560 gagatggagccacacatgagtttgcacacggcctactgcagcaacttgtaccagaatct 4619
 Db 5526 gagatggagccacacatgagtttgcacacggcctactgcagcaacttgtaccagaatct 5585
 QY 4620 catgtccacaaaccccatgtaaaactttcaaccactcaaaagctgtttatttcggctgaagaa 4679
 Db 5586 catgtccacaaaccccatgtaaaactttcaaccactcaaaagctgtttatttcggctgaagaa 5645
 QY 4680 ataacttttttctcacccagtcatttgcacctcttcataatggctgtgtcgcacccctcc 4739

Db 5646 ataacttttttctcacccagtcatttgcacctcttcataatggctgtgtcgcacccctcc 5705
 QY 4740 agaaacgtggttatacttccagtcagtcgtggagaacactgaagactccggttggctcagg 4799
 Db 5706 agaaacgtggttatacttccagtcagtcgtggagaacactgaagactccggttggctcagg 5765
 QY 4800 aactgaggggtgaccttcgggaaggaagtccactcacttatttattatgcctgtgatg 4859
 Db 5766 aactgaggggtgaccttcgggaaggaagtccactcacttatttattatgcctgtgatg 5825
 QY 4860 tgggtcctgcagggagacatccagtcactcgtgtgttctttaaagtcacacctgggaactcgt 4919
 Db 5826 tgggtcctgcagggagacatccagtcactcgtgtgttctttaaagtcacacctgggaactcgt 5885
 QY 4920 gtttattggcctctcttttggggcaatccctcgttttggatgaagtcgagggaatacacagagta 4979
 Db 5886 gtttattggcctctcttttggggcaatccctcgttttggatgaagtcgagggaatacacagagta 5945
 QY 4980 aaagaaattgtctccacccctgaagcgggagtcctcgtctcacttctggaatgtgtcag 5039
 Db 5946 aaagaaattgtctccacccctgaagcgggagtcctcgtctcacttctggaatgtgtcag 6005
 QY 5040 ccactggggacagttctgccccgggcaatggtgtttctcaagtcctctcaaatataatc 5099
 Db 6006 ccactggggacagttctgccccgggcaatggtgtttctcaagtcctctcaaatataatc 6065
 QY 5100 cctattcttacaataactctggccctgagatggtttttaagaagaactcctgtgtcccatgg 5159
 Db 6066 cctattcttacaataactctggccctgagatggtttttaagaagaactcctgtgtccmatgg 6125
 QY 5160 tctccacactcaccatcacctcgtctgtagcaagagtcctcagtcaggggaggtgcatttt 5219
 Db 6126 tctccacactcaccatcacctcgtctgtagcaagagtcctcagtcaggggaggtgcatttt 6185
 QY 5220 agtagttaattgcactttaaccatgagataaaataaaaggaggaactgtttttatcagtcgga 5279
 Db 6186 agtagttaattgcactttaaccatgagataaaataaaaggaggaactgtttttatcagtcgga 6245
 QY 5280 ggttaacactaaaatttcaaaagtgcctcttttgaatactttggcctctctctctgtaga 5339
 Db 6246 ggttaacactaaaatttcaaaagtgcctcttttgaatactttggcctctctctctgtaga 6305
 QY 5340 accaaatggcccttctgtgcacggcctgcacactgaactgagagttctcagctccctgca 5399
 Db 6306 accaaatggcccttctgtgcacggcctgcacactgaactgagagttctcagctccctgca 6365
 QY 5400 gctcactgagccacagactaggtcttctggctcctctccgc 5441
 Db 6366 gctcactgagccacagactaggtcttctggctcctctccgc 6407

RESULT 2
 AAV81825
 ID AAV81825 standard; cDNA; 6407 BP.
 XX
 AAV81825;
 AC
 XX
 DT 11-MAR-1999 (first entry)
 XX
 DE Human rchd528 encoding cDNA.
 KW Human; cardiovascular disease; atherosclerosis; ischaemia; restenosis;
 KW reperfusion; hypertension; arterial inflammation; diagnosis; rchd528; ds.
 OS Homo sapiens.
 FH Key
 FT CDS 19..4464
 FT /*tag= a
 XX
 PN US5849578-A.
 XX

[illegible]

QY 1380 gctctgcccctctgttcaaaatcccaccattttatttttcatcaattttaccatcaaacag 1439
 Db 2346 gctctgcccctctgttcaaaatcccaccattttatttttcaattttaccatcaaacag 2405
 QY 1440 ggcctctgtgcatactactaaagtctactctgtgatgcacacccaatgggtcttctccacc 1499
 Db 2406 ggcctctgtgcatactactaaagtctactctgtgatgcacacccaatgggtcttctccacc 2465
 QY 1500 atcacctttaccagatcccttaacagacatatcatcttgcctccactttctgtctcaaaaac 1559
 Db 2466 atcacctttaccagatcccttaacagacatatcatcttgcctccactttctgtctcaaaaac 2525
 QY 1560 aacettgcccacagtcattcttctacccctgtctcgcccaggccaaggagagactcctgtgac 1619
 Db 2526 aacettgcccacagtcattcttaccctgtctcgcccaggccaaggagagactcctgtgac 2585
 QY 1620 ttaatttcaagacatacaaatgacatcatctcatgacaatgtctccatagtgtcaaaactgc 1679
 Db 2586 ttaatttcaagacatacaaatgacatcatctcatgacaatgtctccatagtgtcaaaactgc 2645
 QY 1680 agaccttaagagccagagcaccacacccaagagagaagtctattacagaaatcaaaagtcaac 1739
 Db 2646 agaccttaagagccagagcaccacacccaagagagaagtctattacagaaatcaaaagtcaac 2705
 QY 1740 aagcctgggtctctgcccacagagtcacaccaaagctgttaacaacaactctcctttggc 1799
 Db 2706 aagcctgggtctctgcccacagagtcacaccaaagctgttaacaacaactctcctttggc 2765
 QY 1800 tcaatccttaacagagtccttccacagagcaaaaccccttccaggccaacagcccaacttagc 1859
 Db 2766 tcaatccttaacagagtccttccacagagcaaaaccccttccaggccaacagcccaacttagc 2825
 QY 1860 acaaaatgtctccaaacttcaaaactaccattctgaaagacctctcagcctcttatgaccac 1919
 Db 2826 acaaaatgtctccaaacttcaaaactaccattctgaaagacctctcagcctcttatgaccac 2885
 QY 1920 tcttggcaccctgtcaagcagacgatctctgttgcactggccctatagccgtacagactac 1979
 Db 2886 tcttggcaccctgtcaagcagacgatctctgttgcactggccctatagccgtacagactac 2945
 QY 1980 agctggaaaacagctctcgtgacccatcctgaaatactgatttctcctcaaatctcaacaga 2039
 Db 2946 agctggaaaacagctctcgtgacccatcctgaaatactgatttctcctcaaatctcaacaga 3005
 QY 2040 aggtggcatcagcacagaaaggaaccgagtgattgtgagtgctaccactggattgattccc 2099
 Db 3006 aggtggcatcagcacagaaaggaaccgagtgattgtgagtgctaccactggattgattccc 3065
 QY 2100 tttagcagtgctaccacatacagcaaaagaaatgaccacaagaacttggtgctttacagcaga 2159
 Db 3066 tttagcagtgctaccacatacagcaaaagaaatgaccacaagaacttggtgctttacagcaga 3125
 QY 2160 gtacagcccagcttcacgttccctcgggaacatactccttctcccaaacacacagttgtttc 2219
 Db 3126 gtacagcccagcttcacgttccctcgggaacatactccttctcccaaacacacagttgtttc 3185
 QY 2220 cacggctgaagacttggctccccaatctgcacacttctgtgttcagagcagcacacagtc 2279
 Db 3186 cacggctgaagacttggctccccaatctgcacacttctgtgttcagagcagcacacagtc 3245
 QY 2280 accaaacaaactgtcctcttcaagcctcagtcacacagctgtgtgtgaaccttgtcttca 2339
 Db 3246 accaaacaaactgtcctcttcaagcctcagtcacacagctgtgtgtgaaccttgtcttca 3305
 QY 2340 caatggcggaatgcgtgcagacaacacagccgtggctaccactgaggtgcccgccttc 2399
 Db 3306 caatggcggaatgcgtgcagacaacacagccgtggctaccactgaggtgcccgccttc 3365
 QY 2400 ctggcaagggaatgatgcagtggtggaatggaatgagctgcgtgcgaacccctgcccactc 2459
 Db 3366 ctggcaagggaatgatgcagtggtggaatggaatgagctgcgtgcgaacccctgcccactc 3425
 QY 2460 cacagccacgtgcaacaataactcagggatccttattctgcaaatgcccgggtgggtacca 2519

Db 3426 cacagccacgtgcacaataactcaggatcctttatctgcaaatgcccgggttgggtacca 3485
 QY 2520 gttggaaaaagggtatgtcaatttgggttagaaccttctgtgacagagtttaaattaaagag 2579
 Db 3486 gttggaaaaagggtatgtcaatttgggttagaaccttctgtgacagagtttaaattaaagag 3545
 QY 2580 aactttcttaatacaactgtgaaaaacattcagacctcagacactcaagaagtgtgaaatgagat 2639
 Db 3546 aactttcttaatacaactgtgaaaaacattcagacctcagacactcaagaagtgtgaaatgagat 3605
 QY 2640 caccaaaacgtttaaataatgtgtttttcagcgtttaccctagtttacatccgatctacagttca 2699
 Db 3606 caccaaaacgtttaaataatgtgtttttcagcgtttaccctagtttacatccgatctacagttca 3665
 QY 2700 cgcctctaggaggatcccaacgcggtgggtgatctcaactgcacaaacacccctttccctggcctc 2759
 Db 3666 cgcctctaggaggatcccaacgcggtgggtgatctcaactgcacaaacacccctttccctggcctc 3725
 QY 2760 caatgtgacgctatttgcactcgtcgtgatagatgcagaaaatgtgtcaactcctcgaagtc 2819
 Db 3726 caatgtgacgctatttgcactcgtcgtgatagatgcagaaaatgtgtcaactcctcgaagtc 3785
 QY 2820 ctctgctgaggtctgcagactccttgggtatctcagaggcgggtcttttagagcgggcagctt 2879
 Db 3786 ctctgctgaggtctgcagactccttgggtatctcagaggcgggtcttttagagcgggcagctt 3845
 QY 2880 gtgcacacgggaagatcccgaaatgtgacaaagacacactcctatctgcactgcactggagcgg 2939
 Db 3846 gtgcacacgggaagatcccgaaatgtgacaaagacacactcctatctgcactgcactggagcgg 3905
 QY 2940 cgttgcctcgtgcagctgcaggtcgggtacttccagtccaacagatgggacctcctcg 2999
 Db 3906 cgttgcctcgtgcagctgcaggtcgggtacttccagtccaacagatgggacctcctcg 3965
 QY 3000 ccgagcatgtgaagatggatataggttgtaaaatgaaacactgcactgaggttggccatttgg 3059
 Db 3966 ccgagcatgtgaagatggatataggttgtaaaatgaaacactgcactgaggttggccatttgg 4025
 QY 3060 ccttgggtggtctcaactgtgaaacccctatcagcttatcactgtgtgtgactgcagcgcgc 3119
 Db 4026 ccttgggtggtctcaactgtgaaacccctatcagcttatcactgtgtgtgactgcagcgcgc 4085
 QY 3120 gggagggtgggtcctcgtcctcctagcactcagcactgattgttaccctgttgcagaaagaa 3179
 Db 4086 gggagggtgggtcctcgtcctcctagcactcagcactgattgttaccctgttgcagaaagaa 4145
 QY 3180 taaaaatgacataaagcaaaactccttcaaaagtggagatttccaaaatgttccccatattgc 3239
 Db 4146 taaaaatgacataaagcaaaactccttcaaaagtggagatttccaaaatgttccccatattgc 4205
 QY 3240 tgaatacccccacaaaatcctcgtccacagaatggggcggagagacttatgaaaatgcata 3299
 Db 4206 tgaatacccccacaaaatcctcgtccacagaatggggcggagagacttatgaaaatgcata 4265
 QY 3300 gaattgggaagtaccacaaaacccctcctccagatgcagcgaatgtgtactactcgcctacaagt 3359
 Db 4266 gaattgggaagtaccacaaaacccctcctccagatgcagcgaatgtgtactactcgcctacaagt 4325
 QY 3360 aaggaaatccagaaacttgaacgaaacggactctaccccggcctacactggactgcagagatc 3419
 Db 4326 aaggaaatccagaaacttgaacgaaacggactctaccccggcctacactggactgcagagatc 4385
 QY 3420 acggcatctctgcattttcccccggacagatataaacccttcttctcatcagttgataagcag 3479
 Db 4386 acggcatctctgcattttcccccggacagatataaacccttcttctcatcagttgataagcag 4445
 QY 3480 aagaagagactacttttaagtccaggagagagagagactcattgtctgagccagtcacc 3539
 Db 4446 aagaagagactacttttaagtccaggagagagagagagactcattgtctgagccagtcacc 4505
 QY 3540 tgggacctctgctcagaggagccgcacacagaggtcgcgccagagatttgcgggagccac 3599

Db 4506 tgggaacctctgtcagagaccgcacacagaggctgcgccaggatttgtcgggagccac 4565
Qy 3600 gctgagtggaagcagagaagaggacaggaatgctggggcggtgacacagtggaagagaca 3659
Db 4566 gctgagtggaagcagagaagaggacaggaatgctggggcggtgacacagtggaagagaca 4625
Qy 3660 ggtgagtggaagcagagcgtctctatcagacaccttctgttactgtgaacgtgaat 3719
Db 4626 ggtgagtggaagcagagcgtctctatcagacaccttctgttactgtgaacgtgaat 4685
Qy 3720 gtgggccaagtatacgaagagctctctctgtgagtgactgacacatggcactggcaccaggcg 3779
Db 4686 gtgggccaagtatacgaagagctctctctgtgagtgactgacacatggcactggcaccaggcg 4745
Qy 3780 actatagccagggcacacacactagaactcagtgacagggaccctgtttcccttcggttg 3839
Db 4746 actatagccagggcacacacactagaactcagtgacagggaccctgtttcccttcggttg 4805
Qy 3840 caacttagtaaatgggtggaggtttcccttttgatctgttttgagactgttccagaaa 3899
Db 4806 caacttagtaaatgggtggaggtttcccttttgatctgttttgagactgttccagaaa 4865
Qy 3900 gaagggttctttcccgagacacttccatagcgacgaatttggtgattcatttgcagcaa 3959
Db 4866 gaagggttctttcccgagacacttccatagcgacgaatttggtgattcatttgcagcaa 4925
Qy 3960 aatactggtgttaattatttctgtccagcgccctgtgtctaaacacagatgagga 4019
Db 4926 aatactggtgttaattatttctgtccagcgccctgtgtctaaacacagatgagga 4985
Qy 4020 tgagcgtaccacgaagctgaagatgcgccattgaacgacagtgatttccatatttt 4079
Db 4986 tgagcgtaccacgaagctgaagatgcgccattgaacgacagtgatttccatatttt 5045
Qy 4080 ctagggtgttattgctacagtttccaaagccagcccccacagtgaggaaatgtgtgaggc 4139
Db 5046 ctagggtgttattgctacagtttccaaagccagcccccacagtgaggaaatgtgtgaggc 5105
Qy 4140 accgcacacaaactgcaatgtgtttttaaagtcgaagtgacacatgtatttaagattttt 4199
Db 5106 accgcacacaaactgcaatgtgtttttaaagtcgaagtgacacatgtatttaagattttt 5165
Qy 4200 tttaaaatctcttgcaagtaaatctcaacttttcaacaagcctggtacagggcgaaac 4259
Db 5166 tttaaaatctcttgcaagtaaatctcaacttttcaacaagcctggtacagggcgaaac 5225
Qy 4260 aacttatattgtgttttagctgaggtcagcagggcagattgagcaggggggcaacttt 4319
Db 5226 aacttatattgtgttttagctgaggtcagcagggcagattgagcaggggggcaacttt 5285
Qy 4320 tcatcatagagggccagcctgggctggagctgacatcagatcaccattgtgagggccagagg 4379
Db 5286 tcatcatagagggccagcctgggctggagctgacatcagatcaccattgtgagggccagagg 5345
Qy 4380 cagctgctatgagggagaaatgtcaaaactgaacgaggtttccaccacttaggaaagca 4439
Db 5346 camctgctatgagggagaaatgtcaaaactgaacgaggtttccaccacttaggaaagca 5405
Qy 4440 gcttgttagccccctgcagctgagtggttggttagaggatgggtgaatagcaggggttaga 4499
Db 5406 gcttgttagccccctgcasctgagtggttggttagaggatgggtgaatagcaggggttaga 5465
Qy 4500 ttctctgcatcaacagtgctttgggaagctgtgttgattcctcaggaagaacaggagcc 4559
Db 5466 ttctctgcatcaacagtgctttgggaagctgtgttgattcctcaggaagaacaggagcc 5525
Qy 4560 gagatggagccacacatgagtttgcacccgctactgcagcactttgtaccaccagaatct 4619
Db 5526 gagatggagccacacatgagtttgcacccgctactgcagcactttgtaccaccagaatct 5585
Qy 4620 catgtccaaaaaccccatgttaaactttcaaccactcaaaactgtttatttcggtgaagaa 4679
Db 5586 catgtccaaaaaccccatgttaaactttcaaccactcaaaactgtttatttcggtgaagaa 5645

Qy 4680 ataacttttttttttcacccaagtcatattgttacctctctatagctgtgtcgcaccctcc 4739
Db 5646 ataacttttttttttcacccaagtcatattgttacctctctatagctgtgtcgcaccctcc 5705
Qy 4740 agaaacgtgggttatacttccagtcagtggtggagaactgaagactccggttgcgaggg 4799
Db 5706 agaaacgtgggttatacttccagtcagtggtggagaactgaagactccggttgcgaggg 5765
Qy 4800 aactgaggggttgaccttcgggaagaagtccactcatcttatttattatccctgtgatg 4859
Db 5766 aactgaggggttgaccttcgggaagaagtccactcatcttatttattatccctgtgatg 5825
Qy 4860 tgggtcctgcagggagacatccagtcactcgtgtcttttaattggcaccctggggaactgt 4919
Db 5826 tgggtcctgcagggagacatccagtcactcgtgtctttaaattggcaccctggggaactgt 5885
Qy 4920 gtttattggccttcttctggggcattcctcctgttttggaagtgaggggaacacagagta 4979
Db 5886 gtttattggccttcttctggggcattcctcctgttttggaagtgaggggaacacagagta 5945
Qy 4980 aaagaattgtctcacacctgaagcggggagtcctcctcatttctggaatgggtgcag 5039
Db 5946 aaagaattgtctcacacctgaagcggggagtcctcctcatttctggaatgggtgcag 6005
Qy 5040 ccaactggggcagttctgcctcggggcagtggttttctcaagtgctctctataataatc 5099
Db 6006 ccaactggggcagttctgcctcggggcagtggttttctcaagtgctctctataataatc 6065
Qy 5100 cctattctacataactccttggccctgatggttttaagcaagaactcctgtgtcccatgg 5159
Db 6066 cctattctacataactccttggccctgatggttttaagcaagaactcctgtgtcccatgg 6125
Qy 5160 tctccacactcaccatcacctcgtctgtagcaagtcctagtcaggggaggtgcatttt 5219
Db 6126 tctccacactcaccatcacctcgtctgtagcaagtcctagtcaggggaggtgcatttt 6185
Qy 5220 agtagttaaattgcaattatccatgagataaaataaaggagaactgttttatcagtgga 5279
Db 6186 agtagttaaattgcaattatccatgagataaaataaaggagaactgttttatcagtgga 6245
Qy 5280 ggtcaactaaaaattcaaaagtgtcgcctttttgaaactcttggcctctctctctagga 5339
Db 6246 ggtcaactaaaaattcaaaagtgtcgcctttttgaaactcttggcctctctctctagga 6305
Qy 5340 accaatggcctttgtgtgcacggcctcgcacctaactgagagattctgagctcctgca 5399
Db 6306 accaatggcctttgtgtgcacggcctcgcacctaactgagagattctgagctcctgca 6365
Qy 5400 gctcactgagccacacagactaggcttcttgctccttcgcg 5441
Db 6366 gctcactgagccacacagactaggcttcttgctccttcgcg 6407

RESULT 3

AA88578
ID AAA88578 standard; cDNA; 6407 BP.
XX
AC AAA88578;
XX
DT 05-FEB-2001 (first entry)
XX
DE Human rchd528 cDNA.
XX
KW Human; rchd528 gene; differential expression; HUVEC; shear stress;
KW endothelial cell; cardiovascular disease; inflammation;
KW atherosclerosis; antiinflammatory; antiatherosclerotic; diagnosis;
KW gene therapy; Ss.
XX
OS Homo sapiens.
XX
FH Key
FT CDS Location/Qualifiers
19..4464

Db 4146 taaaaatgacataagcaaaactatcttccaaaagtggagatttccaaaatgtccccatatgc 4205
Qy 3240 tgaatacccaaaaaactctcgctcacagaatgggcccgaagaactattgaatatgcata 3299
Db 4206 tgaatacccaaaaaactctcgctcacagaatgggcccgaagaactattgaatatgcata 4205
Qy 3300 gaa tgggaagtaccaaaaaactctcccgatgacggatgtgtactactcgcctacaaagt 3359
Db 4266 gaa tgggaagtaccaaaaaactctcccgatgacggatgtgtactactcgcctacaaagt 4325
Qy 3360 aaggaatccagaaacttgaagaaacgactactacccggcctacactggactgcccagatc 3419
Db 4326 aaggaatccagaaacttgaagaaacgactactacccggcctacactggactgcccagatc 4385
Qy 3420 acggcatcttgcatcttcccccagacagtataaacccgtcttctcatcagtgatgaaacag 3479
Db 4386 acggcatcttgcatcttcccccagacagtataaacccgtcttctcatcagtgatgaaacag 4445
Qy 3480 aagaagagactacttttaagtccaggagagagaggagactcattgctctgagccagtacc 3539
Db 4446 aagaagagactacttttaagtccaggagagagaggagactcattgctctgagccagtacc 4505
Qy 3540 tgggaactctgctcagaggaccgacacagaggctgcccagagatttgcggagaccac 3599
Db 4506 tgggaactctgctcagaggaccgacacagaggctgcccagagatttgcggagaccac 4565
Qy 3600 gctgagtggcaagcaggaagagacagggcatgcggggcggtgaccacagtggaagagaca 3659
Db 4566 gctgagtggcaagcaggaagagacagggcatgcggggcggtgaccacagtggaagagaca 4625
Qy 3660 ggtggatggtggauccacacaggctgctcatctcagcaaccttggttactgtgaacgtgaat 3719
Db 4626 ggtggatggtggauccacacaggctgctcatctcagcaaccttggttactgtgaacgtgaat 4685
Qy 3720 gtgggcagatataagagagtctctctgagtactgcacacatggcactggcaccaggcg 3779
Db 4686 gtgggcagatataagagagtctctctgagtactgcacacatggcactggcaccaggcg 4745
Qy 3780 actattagccaggcagaccactagacttcagtgccaggagcctggtttcccttcgtttg 3839
Db 4746 actattagccaggcagaccactagacttcagtgccaggagcctggtttcccttcgtttg 4805
Qy 3840 caactttagtaaaattgggtgggaggtttcccttttggatctgttttgagactggttccagaaa 3899
Db 4806 caactttagtaaaattgggtgggaggtttcccttttggatctgttttgagactggttccagaaa 4865
Qy 3900 gaaggctctctcccgagacacttccataggaagaatttgggtgattcatttgcagcaa 3959
Db 4866 gaaggctctctcccgagacacttccataggaagaatttgggtgattcatttgcagcaa 4925
Qy 3960 aatactggctgttaattttctctgcccagcgctcggtgcttaaacacacagatgagga 4019
Db 4926 aatactggctgttaattttctctgcccagcgctcggtgcttaaacacacagatgagga 4985
Qy 4020 tgaagcgtaccactgaagctctgaagatgtgccattgaaacggacagtgttttcatatgttt 4079
Db 4986 tgaagcgtaccactgaagctctgaagatgtgccattgaaacggacagtgttttcatatgttt 5045
Qy 4080 ctagggtgtcttatgtctacagtttccaaagccagcccccacacagtgaggaatgttgaggc 4139
Db 5046 ctagggtgtcttatgtctacagtttccaaagccagcccccacacagtgaggaatgttgaggc 5105
Qy 4140 acgcacacaaactgcaatgtgtttttaaagtcgaaggtgacacatgtatttaagatttttt 4199
Db 5106 acgcacacaaactgcaatgtgtttttaaagtcgaaggtgacacatgtatttaagatttttt 5165
Qy 4200 tttaaaaactctctgaggttaaaactctcaacttttccaaacagcctggatcaggggcaaac 4259
Db 5166 tttaaaaactctctgaggttaaaactctcaacttttccaaacagcctggatcaggggcaaac 5225
Qy 4260 aacttatatttgggttttagctggaaggtcagcaggcagatgtgaagcaggggggcacttt 4319
Db 5226 aacttatatttgggttttagctggaaggtcagcaggcagatgtgaagcaggggggcacttt 5285

Qy 4320 tcaccatgagggcccagcctggggcctgggactctgatccacttgtcggaggccagagg 4379
Db 5286 tcaccatgagggcccagcctggggcctgggactctgatccacttgtcggaggccagagg 5345
Qy 4380 cagctgcgta tgggaggaaaaatgtcaaaactgaacgcagggtttcaccactctagaaaaa 4439
Db 5346 camctgcgta tgggaggaaaaatgtcaaaactgaacgcagggtttcaccactctagaaaaa 5405
Qy 4440 gcttgttgagcccctgcagctggatgtggttagagggatggctgaaatagcaggttaga 4499
Db 5406 gcttgttgagcccctgcascctggatgtggttagagggatggctgaaatagcaggttaga 5465
Qy 4500 tttcctgcatacaacagtgcttttgggaagctgttggaattcctgaggaagaacacaggagcc 4559
Db 5466 tttcctgcatacaacagtgcttttgggaasctgttggaattcctgaggaagaacacaggagcc 5525
Qy 4560 gagatgagccacacatgagtttgcctcaccggctactgcagcactttgtaccccgaatct 4619
Db 5526 gagatgagccacacatgaaattgctcacggctactgcagcactttgtaccccgaatct 5585
Qy 4620 catctccacaaaaccccatgtaaaactttcaacacactcaaaagctgtttattcggctgaa 4679
Db 5586 catgtccacaaaaccccatgtaaaactttcaacacactcaaaagctgtttattcggctgaa 5645
Qy 4680 ataacttttttctcaccocagtcatttgtacctcttcataatggctgtgcgacccctcc 4739
Db 5646 ataacttttttctcaccocagtcatttgtacctcttcataatggctgtgcgacccctcc 5705
Qy 4740 agaaactgggtta tacttccagtcagtggtgggagaactgaagaactccogtggctgagg 4799
Db 5706 agaaactgggtta tacttccagtcagtggtgggagaactgaagaactccogtggctgagg 5765
Qy 4800 aactgaggggttgaccttcgggaagaagttccactcatcttatttattatcctgtgatg 4859
Db 5766 aactgaggggttgaccttcgggaagaagttccactcatctatttattatcctgtgatg 5825
Qy 4860 tgggtcctgcaggagagacatccagtcactcgtgtcttctaattgcccactctgggaactgt 4919
Db 5826 tgggtcctgcaggagagacatccagtcactcgtgtcttctaattgcccactctgggaactgt 5885
Qy 4920 gtttattggcctcttcttggggcatctcctgggttttggatgaagtgggggaaatcacagagta 4979
Db 5886 gtttattggcctcttcttggggcatctcctgggttttggatgaagtgggggaaatcacagagta 5945
Qy 4980 aaagaattgtctccacactgaacggggagtcgcccttcacattcttggaatggtgcag 5039
Db 5946 aaagaattgtctccacactgaacggggagtcgcccttcacattcttggaatggtgcag 6005
Qy 5040 ccactggggacagttctgcccgggcatggttgttttccaaggtccctctaaataataatc 5099
Db 6006 ccactggggacagttctgcccgggcatggttgttttccaaggtccctctaaataataatc 6065
Qy 5100 cctatttaccataaactcttggccctgatgggtttttaagcaagaactcctgtgtcccatgg 5159
Db 6066 cctatttaccataaactcttggccctgatgggtttttaagcaagaactcctgtgtcccatgg 6125
Qy 5160 tctccacactcacacacactcctgtatgcaagagtccttagtcagggaggtgcatttt 5219
Db 6126 tctccacactcacacacactcctgtatgcaagagtccttagtcagggaggtgcatttt 6185
Qy 5220 agtagtttaaaattgacattatccatgagataaaataaaaggagaactgtttttatcagtgga 5279
Db 6186 agtagtttaacattgcaattatccatgagataaaataaaaggagaactgtttttatcagtgga 6245
Qy 5280 ggcataacctaaaaattcaaaagtgcgctttttgaaatcttgggcctctctctctctgtaga 5339
Db 6246 ggcataacctaaaaattcaaaagtgcgctttttgaaatcttgggcctctctctctctgtaga 6305
Qy 5340 accaaatggccctttgtaggctcacggcctcgcaactaactggagagttcttgagctcctgca 5399
Db 6306 accaaatggccctttgtggtcacaogcctcgcaactaactggagagttcttgagctcctgca 6365

QY 5400 gctcacctgagccacagactaggtcttctggtctcttcctccgc 5441
|||||
Db 6366 gctcacctgagccacagactaggtcttctggtctcttcctccgc 6407
RESULT 4
AA289796
ID AA289796 standard; cDNA; 6407 BP.
XX
AC AA289796;
XX
DT 05-MAY-2000 (first entry)
XX
DE Human cardiovascular disease associated gene rchd528.
XX
KW Differentially expressed; cardiovascular disease; atherosclerosis;
KW ischaemia; reperfusion; hypertension; restenosis; arterial inflammation;
KW rchd528; transmembrane protein; ss.
XX
OS Homo sapiens.
XX
PN US6020463-A.
XX
PD 01-FEB-2000.
XX
PF 06-OCT-1997; 97US-0944423.
XX
PR 09-FEB-1996; 96US-0599654.
PR 10-FEB-1995; 95US-0368844.
PR 07-JUN-1995; 95US-0485573.
XX
XX (BGHM) BRIGHAM & WOMENS HOSPITAL.
PA (MILL-) MILLENNIUM PHARM INC.
XX
XX Glabrone MA, Falb DA;
PI
XX
XX WPI; 2000-146911/13.
DR P-PSDB; AAY78508.
XX
XX
PT Marker proteins for the diagnosis of cardiovascular diseases such as
PT atherosclerosis and hypertension, comprising peptide sequences derived
PT from the rchd523 transmembrane protein -
XX
XX Examples; Fig 30; 121pp; English.
XX
CC This sequence represents the human rchd528 gene sequence. This sequence
CC is related to the rchd523 transmembrane polypeptide encoded by cDNA
CC contained in the plasmid pfchd523. The rchd523 protein is differentially
CC expressed in diseased cells compared to healthy cells. The rchd523
CC protein may be used as a marker protein for the diagnosis of
CC cardiovascular diseases including atherosclerosis, ischaemia,
CC reperfusion, hypertension, restenosis and arterial inflammation. rchd523
CC peptides may be used as antigens in the production of antibodies specific
CC for rchd523. The anti-rchd523 antibodies may then be used in diagnostic
CC assays to quantitate rchd523 peptides in samples.
XX
SQ Sequence 6407 BP; 1594 A; 1762 C; 1520 G; 1509 T; 22 other;

Query Match 73.3%; Score 5098.4; DB 21; Length 6407;
Best Local Similarity 94.3%; Pred. No. 0;
Matches 5412; Conservative 21; Mismatches 8; Indels 301; Gaps 2;

QY 1 gttcgatgaagaattgccgttttcaacaaagagtggaacgcctcgagatgggaac 60
|||||
Db 666 gttcgatgaagaattgccgttttcaacaaagagtggaacgcctcgagatgggaac 725
QY 61 agagagggcagtggtgtcagaagaatggactgtgcacagcgaagggccaccattc 120
|||||
Db 726 agagagggcagtggtgtcagaagaatggactgtgcacagcgaagggccaccattc 785
QY 121 ggcttgagcccttccttctctctgcttggagatggagagctgaccagcctcttag 180
|||||

Db 786 ggcttgagccgctcttctctctggttggagatggagagctgaccaagccttctag 845
QY 181 gaagagaaattctctagaccagatctctctggtcatttctacagagacagcttc 240
|||||
Db 846 gaagagaaattctctagaccagatctctctggtcatttctacagagacagcttc 905
QY 241 ctctctctcttagacatttctcacccttgaagatgagagagcttaacaactccac 300
|||||
Db 906 ctctctctcttagacatttctcacccttgaagatgagagagcttaacaactccac 965
QY 301 tggcctccagagctctcagtcagtcacaacaaagacaatgcattgtgtcacgtgttcac 360
|||||
Db 966 tggcctccagagctctcagtcagtcacaacaaagacaatgcattgtgtcacgtgttcac 1025
QY 361 tgatgtgtgcccagagacgctgcgtatctttgacggtcagctggtgacacctgtgagaacac 420
|||||
Db 1026 tgatgtgtgcccagagacgctgcgtatctttgacggtcagctggtgacacctgtgagaacac 1085
QY 421 agaaggtctcccaagagctcagaattgcccagacttcacctcagctctcttcacc 480
|||||
Db 1086 agaaggtctcccaagagctcagaattgcccagacttcacctcagctctcttcacc 1145
QY 481 ctctgcagtggaaatcgagaagaacagtagagtaactgggaatccaggggatgaggaatt 540
|||||
Db 1146 ctctgcagtggaaatcgagaagaacagtagagtaactgggaatccaggggatgaggaatt 1205
QY 541 cattgaaccatccacagagaaatgaatttggacttacctgttgcg-tggcaaaatgattc 599
|||||
Db 1206 cattgaaccatccacagagaaatgaatttggacttacctgttgcg-tggcaaaatgattc 1265
QY 600 cccaaccttggagaacatcagcttgcagcagctctgagtgcaaaatgggaagctcccat 659
|||||
Db 1266 cccaaccttggagaacatcagcttgcagcagctctgagtgcaaaatgggaagctcccat 1325
QY 660 gtctcagactgagactgtgtctagctcagtcgacccatcagaggtgagagatcactgc 719
|||||
Db 1326 gtctcagactgagactgtgtctagctcagtcgacccatcagaggtgagagatcactgc 1385
QY 720 aacttggtcttgaccacagacacacacatctgcagatgtgacaggaagctctgcttcata 779
|||||
Db 1386 aacttggtcttgaccacagacacacacatctgcagatgtgacaggaagctctgcttcata 1445
QY 780 tctgaaagtgtaagtgttcagttgacccagttctcagacttactgtacagctctgg 839
|||||
Db 1446 tctgaaagtgtaagtgttcagttgacccagttctcagacttactgtacagctctgg 1505
QY 840 aggaagtccacagacgattggagatagagattattcagagttctcatctctctctc 899
|||||
Db 1506 aggaagtccacagacgattggagatagagattattcagagttctcatctctctctc 1565
QY 900 ggaagcttgaattcattcagcaccacgctggagaacgttcaa----- 940
|||||
Db 1566 ggaagcttgaattcattcagcaccacgctggagaacgttcaa----- 1625
QY 941 ----- 940
Db 1626 gccaggccaagcactagtgacagttccgccaatgcagagagacaggaacttctgggtgcc 1685
QY 941 ----- 940
Db 1686 ctctctcgccaccacacaccttggctactgtcactggaaacgggggaacgcacactcggtc 1745
QY 941 ----- 940
Db 1746 tgtcacctcccaacacacagcatgacagcacttctgtgggaagcaggcagccctgcagc 1805
QY 941 ----- 940
Db 1806 ggccatgcccccaagaacacagaggggtgcctctctgcacgtaaacgtgacggagacatggg 1865
QY 941 -----tcgctgggattagctacgg 959
|||||
Db 1866 cctggtctcaggtcacttgcgcctccagctgcactgcaggtgcgtggattagctacgg 1925

QY 960 tcaagtgcgtggcacagctattgaacaaaggactccagcgaccacacagaccacacta 1019
DB 1926 tcaagtgcgtggcacagctattgaacaaaggactccagcgaccacacagaccacacta 1985
QY 1020 cctgtcatctacttccaccaagagaaacggggcttactgtccattacagataacagttc 1079
DB 1986 cctgtcatctacttccaccaagagaaacggggcttactgtccattacagataacagttc 2045
QY 1080 atcctcagacaattgtggagagctcaacttcttatattataataatctcaactcttcacattc 1139
DB 2046 atcctcagacaattgtggagagctcaacttcttatattataatctcaactcttcacattc 2105
QY 1140 agagtattctccctttctcatgctcagactcgagagagaaagtaaacatctcatctctatgacgg 1199
DB 2106 agagtattctccctttctcatgctcagactcgagagagaaagtaaacatctcatctctatgacgg 2165
QY 1200 ggaatatgctcagccttctactgagtcgagtcgagttctgtcatataccaaaccttcogtctcta 1259
DB 2166 ggaatatgctcagccttctactgagtcgagtcgagttctgtcatataccaaaccttcogtctcta 2225
QY 1260 cacaccacacaattaatatgcgaacaaacttcgggtgttcttggaactgatgctgagtttgt 1319
DB 2226 cacaccacacaattaatatgcgaacaaacttcgggtgttcttggaactgatgctgagtttgt 2285
QY 1320 tagtgactcctcctcctctctcctcctcctcctcctcctcctcctcctcctcctcctcctt 1379
DB 2286 tagtgactcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctcctt 2345
QY 1380 gccctcgcctctgtgtcaaatccccaccatttatatttcatcaattttacatacaacag 1439
DB 2346 gccctcgcctctgtgtcaaatccccaccatttatatttcatcaattttacatacaacag 2405
QY 1440 ggctctgtgcatactactaaagtctactcctgatgcataccacacacttggtctcctcacc 1499
DB 2406 ggctctgtgcatactactaaagtctactcctgatgcataccacacacttggtctcctcacc 2465
QY 1500 atcacctttaccagtaaccttaacgacatactacatctgccccactttctgtctcacaaac 1559
DB 2466 atcacctttaccagtaaccttaacgacatactacatctgccccactttctgtctcacaaac 2525
QY 1560 aaccttgccacagtcaattcttaccctcgttccctggcagggcgaaggagactcctgtgac 1619
DB 2526 aaccttgccacagtcaattcttaccctcgttccctggcagggcgaaggagactcctgtgac 2585
QY 1620 ttcatttcagacatacaaatgacatactcatgacaaatgctccatgtagtcaaaactgc 1679
DB 2586 ttcatttcagacatacaaatgacatactcatgacaaatgctccatgtagtcaaaactgc 2645
QY 1680 agaccttaagagccagagcaccacacacccaagagaagtgattattacagaaatcaaaagtcaac 1739
DB 2646 agaccttaagagccagagcaccacacacccaagagaagtgattattacagaaatcaaaagtcaac 2705
QY 1740 aagccttggtgtctgtcccacagagtcacacaaagctgtataacaaacactctccttggcc 1799
DB 2706 aagccttggtgtctgtcccacagagtcacacaaagctgtataacaaacactctccttggcc 2765
QY 1800 tccatccttaacagagtcctccacagagcaaaccttccagccacagacacacacttagc 1859
DB 2766 tccatccttaacagagtcctccacagagcaaaccttccagccacagacacacacttagc 2825
QY 1860 acaaatgtctccaaacttccaaactacatacttgaaagacctctcagacctcttatgaccac 1919
DB 2826 acaaatgtctccaaacttccaaactacatacttgaaagacctctcagacctcttatgaccac 2885
QY 1920 tctcggcacacctgtcaagacagacatactctggtcaactggccctatagccgtacagactac 1979
DB 2886 tctcggcacacctgtcaagacagacatactctggtcaactggccctatagccgtacagactac 2945
QY 1980 agctggaaaaacagctctcgtcgtgacccatcctgaaataactagttctcctcaaatctcaacaga 2039
DB 2946 agctggaaaaacagctctcgtcgtgacccatcctgaaataactagttctcctcaaatctcaacaga 3005

QY 2040 agtgggcatcagcacagaaagaaacccagtgattgtggtactaccactggattgatccc 2099
DB 3006 agtgggcatcagcacagaaagaaacccagtgattgtggtactaccactggattgatccc 3065
QY 2100 tttagccagtgtagcccatcagcaaaagaaatgaccacaaagcttggcgtttacagcaga 2159
DB 3066 tttagccagtgtagcccatcagcaaaagaaatgaccacaaagcttggcgtttacagcaga 3125
QY 2160 gtacagcccagcttcaagttccctcgggaacatctccttcccaaacacagttgttttc 2219
DB 3126 gtacagcccagcttcaagttccctcgggaacatctccttcccaaacacagttgttttc 3185
QY 2220 cacggctgaaagacttggctcccacaaatctgccacctttgtctgttcagagcagcacacagtc 2279
DB 3186 cacggctgaaagacttggctcccacaaatctgccacctttgtctgttcagagcagcacacagtc 3245
QY 2280 accaacacactgtcctcttccagcctcagtcacacagctgtgtgtgaacaccttgtcttca 2339
DB 3246 accaacacactgtcctcttccagcctcagtcacacagctgtgtgtgaacaccttgtcttca 3305
QY 2340 caatggcggaatgcgtcgcagacacacccagccgtgtgctacacactgcaggtgccgccttc 2399
DB 3306 caatggcggaatgcgtcgcagacacacccagccgtgtgctacacactgcaggtgccgccttc 3365
QY 2400 ctggcaagggggatgattgcagtgaggatggaatgagtgccgtctgcaacccctgcacctc 2459
DB 3366 ctggcaagggggatgattgcagtgaggatggaatgagtgccgtctgcaacccctgcacctc 3425
QY 2460 cacagccacgtgcacaatactcaggggataccttttatctgcaaaatgcgcggttgggtacca 2519
DB 3426 cacagccacgtgcacaatactcaggggataccttttatctgcaaaatgcgcggttgggtacca 3485
QY 2520 gttggaaaaagggatatgcgaatttgggttagaaccttcgtgcagagtttaattaaaaag 2579
DB 3486 gttggaaaaagggatatgcgaatttgggttagaaccttcgtgcagagtttaattaaaaag 3545
QY 2580 aactttctttaatacaactgtggaaaaacatttcagacctcagacacaaagtgtgaaaaatgagat 2639
DB 3546 aactttctttaatacaactgtggaaaaacatttcagacctcagacacaaagtgtgaaaaatgagat 3605
QY 2640 caccaaaacgtttaaataatgtgttttcaagcgtttacctagttacatccgatctacagtcca 2699
DB 3606 caccaaaacgtttaaataatgtgttttcaagcgtttacctagttacatccgatctacagtcca 3665
QY 2700 cgctctagggagtcaccaacgcggtgggtgactcactgcacaaacacttttccctggcctc 2759
DB 3666 cgctctagggagtcaccaacgcggtgggtgactcactgcacaaacacttttccctggcctc 3725
QY 2760 caatgtgacgctatttgacctgggctgataggatgcagaaaatgtgtcaactcctcgcaagtc 2819
DB 3726 caatgtgacgctatttgacctgggctgataggatgcagaaaatgtgtcaactcctcgcaagtc 3785
QY 2820 ctctcgtgaggtctcgcagctccttgggactcagagggcggtattcttagagcggcagctt 2879
DB 3786 ctctcgtgaggtctcgcagctccttgggactcagagggcggtattcttagagcggcagctt 3845
QY 2880 gtgcgaagcgggaaggtcccgaatgtgacaaagacacctcccctctgcactgcacctggacgg 2939
DB 3846 gtgcgaagcgggaaggtcccgaatgtgacaaagacacctcccctctgcactgcacctggacgg 3905
QY 2940 cgttgcacctgtgcagtgcaagtgcgggatacttctgaatgtcaacaagaatggaccactcctg 2999
DB 3906 cgttgcacctgtgcagtgcaagtgcgggatacttctgaatgtcaacaagaatggaccactcctg 3965
QY 3000 ccgagcattgtgaagatggataggcttgaataaataaaccctgcattgattgccatttgg 3059
DB 3966 ccgagcattgtgaagatggataggcttgaataaataaaccctgcattgattgccatttgg 4025
QY 3060 ccttgggtggtctcaactgtgaaacccctatcagcttatcactgtggtgactgcagcgcgc 3119
DB 4026 ccttgggtggtctcaactgtgaaacccctatcagcttatcactgtggtgactgcagcgcgc 4085
QY 3120 gggagggtgggctcctgtctcatccttaggcatacctgactgattgttaccctgttcgagaaaaa 3179

QY 941 -----tcgctgggattagtaagg 959
Db 1866 cctgggtctcagg tcaatggccgctccagtgcaactcggagtgctgggattagctcagg 1925
QY 960 tcaagtgcgtggcacagctattgaacaaagacttccagcgaccacagaccacacta 1019
Db 1926 tcaagtgcgtggcacagctattgaacaaagacttccagcgaccacagaccacacta 1985
QY 1020 cctgtcatctacttccacaaagagaaacggggttactgtccattacagataacagttc 1079
Db 1986 cctgtcatctacttccacaaagagaaacggggttactgtccattacagataacagttc 2045
QY 1080 atcctcagacatgtgagagctcaactcttattataaatactcaaaccttccacattc 1139
Db 2046 atcctcagacatgtgagagctcaactcttattataaatactcaaaccttccacattc 2105
QY 1140 agagtattctctctctctcatgctcagactgagagaaglaaacatctcatcctatgacgg 1199
Db 2106 agagtattctctctctctcatgctcagactgagagaaglaaacatctcatcctatgacgg 2165
QY 1200 ggaatagtcagccttctactagtcgagtcgagttctgtcatatccaaacttccgctcta 1259
Db 2166 ggaatagtcagccttctactagtcgagtcgagttctgtcatatccaaacttccgctcta 2225
QY 1260 cacaccacacattaatgcccgaacacttcgggttctgttgagacactgatgctgagtttgt 1319
Db 2226 cacaccacacattaatgcccgaacacttcgggttctgttgagacactgatgctgagtttgt 2285
QY 1320 tagtgact 1379
Db 2286 tagtgact 2345
QY 1380 gctctgctctctgtgtccaaatcccaacttattttctcatcaatatttaacatcaaacag 1439
Db 2346 gctctgctctctgtgtccaaatcccaacttattttctcatcaatatttaacatcaaacag 2405
QY 1440 ggcctctgtcatctactaaagtctactctgtatgcatccacacacatggcttctctcacc 1499
Db 2406 ggcctctgtcatctactaaagtctactctgtatgcatccacacatggcttctctcacc 2465
QY 1500 atcacctttaccagtaaccttaagacatctacatctgccccactttctgtctccacaaac 1559
Db 2466 atcacctttaccagtaaccttaagacatctacatctgccccactttctgtctccacaaac 2525
QY 1560 aaccttgccacagtcattcttaccctctgctgcccagggcagggagactcctgtgac 1619
Db 2526 aaccttgccacagtcattcttaccctctgctgcccagggcagggagactcctgtgac 2585
QY 1620 ttcatttcagacatcaacaatgacatcattcatgacaatgctccatagtagtcaaaactgc 1679
Db 2586 ttcatttcagacatcaacaatgacatcattcatgacaatgctccatagtagtcaaaactgc 2645
QY 1680 agaccttaagagcagagcaccacacacaggaagagtagtattacagaatcaaaagtcaacc 1739
Db 2646 agaccttaagagcagagcaccacacacaggaagagtagtattacagaatcaaaagtcaacc 2705
QY 1740 aagcctgggtgtctgtgccccacagtcaccaaaagctgttaacacaaactctctctggcc 1799
Db 2706 aagcctgggtgtctgtgccccacagtcaccaaaagctgttaacacaaactctctctggcc 2765
QY 1800 tccatccttaaacagagtcctccacagagcaaaaccccttccagccacacagcaaccacttagc 1859
Db 2766 tccatccttaaacagagtcctccacagagcaaaaccccttccagccacacagcaaccacttagc 2825
QY 1860 acaaatgtctccaaacttccaaactacattcttgaagacctctcagcctcttatgaccac 1919
Db 2826 acaaatgtctccaaacttccaaactacattcttgaagacctctcagcctcttatgaccac 2885
QY 1920 tcttggaacctgtcaagcacagcatctctgttgcactggcccttatagccgtacagactac 1979
Db 2886 tcttggaacctgtcaagcacagcatctctgttgcactggcccttatagccgtacagactac 2945

QY 1980 agctggaaaaacagctctcgctgaccccatcctgaaatactagttcctcaaatctcaacaga 2039
Db 2946 agctggaaaaacagctctcgctgaccccatcctgaaatactagttcctcaaatctcaacaga 3005
QY 2040 aggtggcatcagcacagaaagaaacccagtgatgtgtgatatcaccactgattgatcc 2099
Db 3006 aggtggcatcagcacagaaagaaacccagtgatgtgtgatatcaccactgattgatcc 3065
QY 2100 tttagccagtgatccaccatcagcaaaagaatagaccacaaagcttggcgtttacagcaga 2159
Db 3066 tttagccagtgatccaccatcagcaaaagaatagaccacaaagcttggcgtttacagcaga 3125
QY 2160 gtacagcccagcttccagttccctcgaaacatctctctcccaaaacacagttgtttc 2219
Db 3126 gtacagcccagcttccagttccctcgaaacatctctctcccaaaacacagttgtttc 3185
QY 2220 cagcgttgaagacttggctcccaaatctgcacactttgtgtttcagagcagcacacagtc 2279
Db 3186 cagcgttgaagacttggctcccaaatctgcacactttgtgtttcagagcagcacacagtc 3245
QY 2280 accaaacacactgtcctcttcagcctcagtcacacagctgtgtgtgaacaccttgtcttca 2339
Db 3246 accaaacacactgtcctcttcagcctcagtcacacagctgtgtgtgaacaccttgtcttca 3305
QY 2340 caatggcggaatgcgtcgacagacaacacagccgtggctaccactgcaggtgccgccttc 2399
Db 3306 caatggcggaatgcgtcgacagacaacacagccgtggctaccactgcaggtgccgccttc 3365
QY 2400 ctggcaaggggagatgtgcagtggtgagtgaatgagtgctgtcgaaacctgcccatc 2459
Db 3366 ctggcaaggggagatgtgcagtggtgagtgaatgagtgctgtcgaaacctgcccatc 3425
QY 2460 cacagccacgtgcacaatactcagggatcctttatctgcaaatgcccgttgggtacca 2519
Db 3426 cacagccacgtgcacaatactcagggatcctttatctgcaaatgcccgttgggtacca 3485
QY 2520 gttgaaaaagggagatgtcaatttgggttagaaccttgcagacagtttaaatataaag 2579
Db 3486 gttgaaaaagggagatgtcaatttgggttagaaccttgcagacagtttaaatataaag 3545
QY 2580 aactttcttaatacaactgtggaacacattcagacctacacaaagttaaagtgaat 2639
Db 3546 aactttcttaatacaactgtggaacacattcagacctacacaaagttaaagtgaat 3605
QY 2640 caccacacgtttaaataatgtgttttccagcgttacctagttacatccgatctacagttca 2699
Db 3606 caccacacgtttaaataatgtgttttccagcgttacctagttacatccgatctacagttca 3665
QY 2700 cgctctaggagtcaccaacgcgtggtgatatctcactcacaacacactttccctggcctc 2759
Db 3666 cgctctaggagtcaccaacgcgtggtgatatctcactcacaacacactttccctggcctc 3725
QY 2760 caatgtcacgtatttgacctggctgtagtagtcagaaaagtgtcaactcctgcaagtc 2819
Db 3726 caatgtcacgtatttgacctggctgtagtagtcagaaaagtgtgtcaactcctgcaagtc 3785
QY 2820 ctctgtgaggtctgcagcgtcttgggatatctcagcggtatcttttagagcggcagctt 2879
Db 3786 ctctgtgaggtctgcagcgtcttgggatatctcagcggtatcttttagagcggcagctt 3845
QY 2880 gtgcaacgggaagagtcgccgaatgtgcaaaagacacactcctatctgcactggaacgg 2939
Db 3846 gtgcaacgggaagagtcgccgaatgtgcaaaagacacactcctatctgcactggaacgg 3905
QY 2940 cgttgccctgtgcccagtcgagtcgggatacttccagtccaacaaagtggaccactcctg 2999
Db 3906 cgttgccctgtgcccagtcgagtcgggatacttccagtccaacaaagtggaccactcctg 3965
QY 3000 ccgagcatgtgaagatggatatagccttgaataatgaacacctgcatgagttgcccattgg 3059
Db 3966 ccgagcatgtgaagatggatatagccttgaataatgaacacctgcatgagttgcccattgg 4025
QY 3060 ccttggtgtctcaactgtggaaacccctatcactgttatcactgtgtgtagtcgagcgcg 3119

Db 4026 ccttgggtctcaactgtgaaacccctatcagcttatcactgtggtgatcgcagcgc 4085
QY 3120 gggagggtggtcctcgtcctcattcagctagcgcagctgattgttacctgttgagaaagaa 3179
Db 4086 gggagggtggtcctcgtcctcattcagctagcgcagctgattgttacctgttgagaaagaa 4145
QY 3180 taaaaatgacataagcaaaactcatttcaaaaagtggagatttccaaaatgtccccatatgc 3239
Db 4146 taaaaatgacataagcaaaactcatttcaaaaagtggagatttccaaaatgtccccatatgc 4205
QY 3240 tgaatacccaaaaaatcctcgtctcaacaagaattggggccgagaaagctattgaaatgcatga 3299
Db 4206 tgaatacccaaaaaatcctcgtctcaacaagaattggggccgagaaagctattgaaatgcatga 4265
QY 3300 gaattgggaagtaccaaaaactcctccagatgcaggtatggttactactcgcctcaaatgt 3359
Db 4266 gaattgggaagtaccaaaaactcctcccaatgacggaagtggagggatgctactcaaatgt 4325
QY 3360 aaggaaatccagaacttgaacgaaacggaactctacccggcctacactggaactgccagatc 3419
Db 4326 aaggaaatccagaacttgaacgaaacggaactctacccggcctacactggaactgccagatc 4385
QY 3420 acggcattcttgaaatttccccggagacgtataaacccgtcttccatcagtgatgaaagcag 3479
Db 4386 acggcattcttgaaatttccccggagacgtataaacccgtcttccatcagtgatgaaagcag 4445
QY 3480 aagaagagactactttaaagtcacagagagagagggagactcatgctcgtgagccagtcacc 3539
Db 4446 aagaagagactactttaaagtcacagagagagagggagactcatgctcgtgagccagtcacc 4505
QY 3540 tgggacctcgtctcagaggaccgacacagagagctgcgcccagatttgcggagagccac 3599
Db 4506 tgggacctcgtctcagaggaccgacacagagagctgcgcccagatttgcggagagccac 4565
QY 3600 gctgagtggaagcaggaaagggagacagcatgcggggcgctgacacagctggaggagaca 3659
Db 4566 gctgagtggaagcaggaaagggagacagcatgcggggcgctgacacagctggaggagaca 4625
QY 3660 ggttggaatggaaccacagctcgtcctcattcagcaccttctgttactgtgaacgtgaat 3719
Db 4626 ggttggaatggaaccacagctcgtcctcattcagcaccttctgttactgtgaacgtgaat 4685
QY 3720 gtgggcccagtatcaagagatctctctgtagctgacactgacacatggcactggcaccagggcg 3779
Db 4686 gtgggcccagtatcaagagatctctctgtagctgacactgacacatggcactggcaccagggcg 4745
QY 3780 actattagccaggcagaccactagacttcagtgacaggagacctgggtttcccttcgtttg 3839
Db 4746 actattagccaggcagaccactagacttcagtgacaggagacctgggtttcccttcgtttg 4805
QY 3840 cactttagtaaaattgggtgggaggtttcccttttgggactcgtttttgagactgtttccagaaa 3899
Db 4806 cactttagtaaaattgggtgggaggtttcccttttgggactcgtttttgagactgtttccagaaa 4865
QY 3900 gaaggcttcctttcccagagacacttccatagggcagcaaatgtgtgattcatttgcagcaa 3959
Db 4866 gaaggcttcctttcccagagacacttccatagggcagcaaatgtgtgattcatttgcagcaa 4925
QY 3960 aatactggctgttaattatttccctgcccagcgctgcgtgctaaacaacagatgagga 4019
Db 4926 aaactggctgtgttaattatttccctgcccagcgcctgcgtgctaaacaacagatgagga 4985
QY 4020 tggagcttaccactgaagctcgaagatgtgccattgaaacggacagtgatttccatatgttt 4079
Db 4986 tggagcttaccactgaagctcgaagatgtgccattgaaacggacagtgatttccatatgttt 5045
QY 4080 ctagggttgtttatgtctacagtttccaagccagcccccacagtgaggaatgtgtgaggc 4139
Db 5046 ctagggttgtttatgtctacagtttccaagccacccccacagtgaggaatgtgtgaggc 5105
QY 4140 accgcacacaaactgcgaatgtgttttttaagtcaaggtgcacacatgtattttaagattttt 4199
,

Db 5106 acgcacacaaactgcaattgtgtttttaagtcaagggtgcacacatgtatttaagattttt 5165
QY 4200 tttaaaaatcctttgcagattaaaaatcctcactttttcaacaagcctggatcagggcaaac 4259
Db 5166 tttaaaaatccttgcagattaaaaatcctcacttttcaacaagcctggatcagggcaaac 5225
QY 4260 aacttatatttgggttttagctggaggctcagcagcgagatgcaggcagggggcacttt 4319
Db 5226 aacttatatttgggttttagctggaggctcagcagcgagatgcaggcagggggcacttt 5285
QY 4320 tcatccatggggggcccagcctggggcctgggaactctcgcacccatttgtggggccagagg 4379
Db 5286 tcatccatggggggcccagcctggggcctgggaactctcgcacccatttgtggggccagagg 5345
QY 4380 cagctgcgtatggagagaaatgtcaaaactgaacgcaggtttccaccactcagaaagca 4439
Db 5346 camctgcgtatggaggggaaatgtcaaaactgaacgcaggtttccaccactcagaaagca 5405
QY 4440 gcttggttgagccctgcagctggatgtgggttagagggtgggctgaaatagcaggttaga 4499
Db 5406 gcttggttgagccctgcacstggatgtgggttagagggtgggctgaaatagcaggttaga 5465
QY 4500 tttcctgcatacaacagctgctttgggaaagctgtgtggattcctggagggaacacagggagcc 4559
Db 5466 tttcctgcatacaacagctgctttgggaaasctgtgtggattcctggagggaacacagggagcc 5525
QY 4560 gagatggagccacacatgagtttgcctcacggctactgcagcacttttgtaccacgaact 4619
Db 5526 gagatggagccacacatgaaattgctcacggctactgcagcacttttgtaccacgaact 5585
QY 4620 catgtccacaaaaccccatgtaaactttcaaacctcacaagctgtttatttcggctgaagaa 4679
Db 5586 catgtccacaaaaccccatgtaaactttcaaacctcacaagctgtttatttcggctgaagaa 5645
QY 4680 ataacttttttttccacccagtcatttgtacctctcctcatatgctgtgtgcgacccctcc 4739
Db 5646 ataacttttttttccacccagtcatttgtacctctcctcatatgctgtgtgcgacccctcc 5705
QY 4740 agaaactgtgtatacttccagtcagtggtgggagaaactgaagactccggttgcgtcgagg 4799
Db 5706 agaaactgtgtatacttccagtcagtggtgggagaaactgaagactccggttgcgtcgagg 5765
QY 4800 aactgaggggttgaccttcgggaaagagttccactcatcttatttattatgcctgtgatg 4859
Db 5766 aactgaggggttgaccttcgggaaagagttccactcatcttatttattatgcctgtgatg 5825
QY 4860 tgggtcctgcaggagagacatccagtaactcgggtgtctttaaattgccaccctgggaaactgt 4919
Db 5826 tgggtcctgcaggagagacatccagtaactcgggtgtctttaaattgccaccctgggaaactgt 5885
QY 4920 gtttattgggcttcttttggggcatcctgggttttggatgaaagtgggggaaacacagaggt 4979
Db 5886 gtttattgggcttcttttggggcatcctgggttttggatgaaagtgggggaaacacagaggt 5945
QY 4980 aaagaaattgtctccacccctgaagcggggagtcgccgcttcacatttcttggaaatgggtcag 5039
Db 5946 aaagaaattgtctccacccctgaagcggggagtcgccgcttcacatttcttggaaatgggtcag 6005
QY 5040 ccactggggagacagttctgcgccgggcaatggttgttttccaaagtcctctaaataataatc 5099
Db 6006 ccactggggagacagttctgcgccgggcaatggttgttttccaaagtcctctaaataataatc 6065
QY 5100 cctattcttacataaactcttggccctgagtggtttttaagcaagaactcctgtgtccactgg 5159
Db 6066 cctattcttacataaactcttggccctgagtggtttttaagcaagaactcctgtgtccmatgg 6125
QY 5160 tctccacccactccactacccctgctgtagcaagagtccttagtcagggggaggtgcatttt 5219
Db 6126 tctccacccactccactacccctgctgtagcaagagtccttagtcagggggaggtgcatttt 6185
QY 5220 agtagttaaatgtcacttatccatgagataaaataaaagggagaaactgtttttatcagtgga 5279
Db 6186 agtagttaaatgtcacttatccatgagataaaataaaagggagavctgtttttatcagtgga 6245

QY 5280 ggctaaactaaatttcaaaagtgcgcctttttgaaatcttgggcctctctctgtaga 5339
 Db 6246 ggctaaactaaatttcaaaagtgcgcctttttgaaatcttgggcctctctctgtaga 6305
 QY 5340 accaatggcccttggctcagcgccctgcacctaactagagagttctgagctccctgca 5399
 Db 6306 accaatggcccttggctcagcgccctgcacctaactagagagttctgagctccctgca 6365
 QY 5400 gctcaactgagccacagactaggtcttcttggctcttcgcgc 5441
 Db 6366 gctcaactgagccacagactaggtcttcttggctcttcgcgc 6407

RESULT 6

AAT36035
 ID AAT36035 standard; DNA; 6407 BP.

AC AAT36035;

XX 20-NOV-1996 (first entry)

DE rchd528 gene differentially expressed in cardiovascular disease.
 XX Cardiovascular disease; differential expression; target gene;
 KW pathway gene; fingerprint gene; atherosclerosis; ischaemia;
 KW reperfusion; hypertension; restenosis; arterial inflammation;
 KW vector; antibody; diagnosis; gene therapy; drug screening;
 KW rchd528 gene; ss.

XX Homo sapiens.

PH Key Location/Qualifiers
 FT CDS 19..4446
 FT /*tag= a

PN W09624604-A1.

XX 15-AUG-1996.

PF 09-FEB-1996; 96WO-U01883.

XX 07-JUN-1995; 95US-0485573.

PR 10-FEB-1995; 95US-0386844.

XX (MILL-) MILLENNIUM PHARM INC.

PA Falb DA;

PI WPI; 1996-384391/38.

XX P-PSDB; AAW03740.

XX New genes differentially expressed in cardiovascular disease - and
 PT related vectors, host cells, proteins and antibodies, for diagnosis,
 PT monitoring, treatment and drug screening

XX Claim 1; Fig 31A-E; 200pp; English.

XX 4 Novel genes, rchd502 (AAT36033), rchd523 (AAT36034), rchd528
 CC (AAT36035) and rchd534 (AAT36036), are differentially expressed in
 CC endothelial cells subjected to shear stress. HUVEC cells were exposed
 CC to laminar shear stress, which is thought to be responsible for the
 CC prevalence of atherosclerotic lesions in areas of unusual
 CC circulatory flow. RNAs from treated and control cells were used to
 CC generate cDNA libraries. Differentially expressed bands were
 CC identified by electrophoresis, subcloned and sequenced. The rchd528
 CC gene product (AAW03740) was also sequenced. rchd528 is very highly
 CC expressed in the heart. 4 other genes (see also AAT36033-36) are
 CC differentially expressed in endothelial cells exposed to interleukin-1.
 CC Detection of these 8 novel genes in excess of normal levels allows the
 CC diagnosis of cardiovascular diseases (CVD). The genes can be used to
 CC generate diagnostic probes, to produce recombinant gene products, to
 CC breed transgenic animal models of CVD, in gene replacement therapy,

CC or (as antisense, ribozyme or triplex sequences) to treat CVD.
 XX Sequence 6407 BP; 1592 A; 1763 C; 1523 G; 1507 T; 22 other;

Query Match 72.5%; Score 5042.4; DB 17; Length 6407;
 Best Local Similarity 93.6%; Pred No. 0;
 Matches 5377; Conservative 21; Mismatches 43; Indels 301; Gaps 2;

QY 1 gttcgtgaaagaattgccgctttttcaacaagaatggaacagcctcggagatgggaac 60
 Db 666 gttcgtgaaagaattgccgctttttcaacaagaatggaacagcctcggagatgggaac 725
 QY 61 agagaggcgatggggtgtcagaagaatggactgtgcacagccaagaggccacacttc 120
 Db 726 agagaggcgatggggtgtcagaagaatggactgtgcacagccaagaggccacacttc 785
 QY 121 ggtctggagcctctcttctctctgttggagatgggagagctgaccacgcctcttag 180
 Db 786 ggtctggagcctctcttctctctgttggagatgggagagctgaccacgcctcttag 845
 QY 181 gaagagaaattctctcaggaccagatctctctgctgctcattctacaggacagagcttc 240
 Db 846 gaagagaaattctctcaggaccagatctctctgctgctcattctacaggacagagcttc 905
 QY 241 ctctctctcttagaccttctctcacttctgaaagtacagagaagctttaaacaactccac 300
 Db 906 ctctctctcttagaccttctctcacttctgaaagtacagagaagctttaaacaactccac 965
 QY 301 tggcctccagagctcctcagtcagtcacaacaagaacatgcatgttctaccgtgttccac 360
 Db 966 tggcctccagagctcctcagtcagtcacaacaagaacatgcatgttctaccgtgttccac 1025
 QY 361 tgatggtggcccgagaacgctggagatcttggacggtcagctctgggacctgtgagcaagac 420
 Db 1026 tgatggtggcccgagaacgctggagatcttggacggtcagctctgggacctgtgagcaagac 1085
 QY 421 agaaaggcttcccaaggactccagaattgccagacttccatctcagctctcttccacc 480
 Db 1086 agaaaggcttcccaaggactccagaattgccagacttccatctcagctctcttccacc 1145
 QY 481 ctctgcagtggaaatcgagaagaacagtagtaactgggaatccaggggagatgggaatt 540
 Db 1146 ctctgcagtggaaatcgagaagaacagtagtaactgggaatccaggggagatgggaatt 1205
 QY 541 cattgaaccatccacagaaatgaatttggacttacgtcttcttggc-tggcaaatgattc 599
 Db 1206 cattgaaccatccacagaaatgaatttggacttacgtcttcttggc-tggcaaatgattc 1265
 QY 600 cccaacctttggagaacatcagcttgcagcagctctgaggtgcaaaatggaagtcccat 659
 Db 1266 cccaacctttggagaacatcagcttgcagcagctctgaggtgcaaaatggaagtcccat 1325
 QY 660 gtctcagactgagactgtgtcagtcagtcgaccctatgagaggtggagagatcaactgc 719
 Db 1326 gtctcagactgagactgtgtcagtcagtcgaccctatgagaggtggagagatcaactgc 1385
 QY 720 aactggtctttagcaacaacagcacaactctgcagatgtgacagagaagctctgttccata 779
 Db 1386 aactggtctttagcaacaacagcacaactctgcagatgtgacagagaagctctgttccata 1445
 QY 780 tcttgaagggtgaatgcttcagttgaccagttctcagactctcagactctgtacagctgg 839
 Db 1446 tcttgaagggtgaatgcttcagttgaccagttctcagactctcagactctgtacagctgg 1505
 QY 840 aggaagtccacacagcattggagatagagttattcagagttcttcacatctacatctctc 899
 Db 1506 aggaagtccacacagcattggagatagagttattcagagttcttcacatctacatctc 1565
 QY 900 ggaagcttgaattcatcagcaccacacgtgagaacgttcaa----- 940
 Db 1566 ggaagcttgaattcatcagcaccacacgtgagaacgttcaa----- 1625

QY 941 ----- 940
Db 1626 gccaggccaagcactaggtgacagttccgcgaatgcagagagacaggacttctgtgggtgcc 1685
QY 941 ----- 940
Db 1686 ctctctcggccaccacaccttggctactgtcactggaacggggaacgcacactcggttc 1745
QY 941 ----- 940
Db 1746 tgtcacccctcaccaacacagatgagcacgacttctgtgggaagcgagccctgcagc 1805
QY 941 ----- 940
Db 1806 ggccatgccccaagaacacagaggtgctctctcgtcacgttaaacgtgacgcagacatggg 1865
QY 941 -----
Db 1866 cctgtgtctcaggtcactggccgctccagtgcactcggagtcgctgggattagctacgg 1925
QY 960 tcaagtgcgtggtgcacagctattgaacaaaagacttccagcgaccacacagacacaccta 1019
Db 1926 tcaagtgcgtggtgcacagctattgaacaaaagacttccagcgaccacacagacacaccta 1985
QY 1020 cctgtcatctactttccaccaaaaggaaacggcggttactgtccattacagataaacagttc 1079
Db 1986 cctgtcatctactttccaccaaaaggaaacggcggttactgtccattacagataaacagttc 2045
QY 1080 atctcagaacattgtggagagctcaacttcttattattaaaaatctcaaaccttccaatc 1139
Db 2046 atctcagaacattgtggagagctcaacttcttattattaaaaatctcaaaccttccaatc 2105
QY 1140 agagtattctctcttctctatctcagactcagagctgagagaagttaacatctcatccttatgacgg 1199
Db 2106 agagtattctctcttctctatctcagactcagagctgagagaagttaacatctcatccttatgacgg 2165
QY 1200 ggaatatgctcagccttctactgagtcgccagttctgcatacatccaaccttccgtccta 1259
Db 2166 ggaatatgctcagccttctactgagtcgccagttctgcatacatccaaccttccgtccta 2225
QY 1260 cacaccacacataataatgcccgaacacttcgggtgttctcgggaactgatctgagtttgt 1319
Db 2226 cacaccacacataataatgcccgaacacttcgggtgttctcgggaactgatctgagtttgt 2285
QY 1320 tagtgaact 1379
Db 2286 tagtgaact 2345
QY 1380 gcctcgtccctctgtgtcaaatcccaaccattttatttttcaatatttaccatacaaccag 1439
Db 2346 gcctcgtccctctgtgtcaaatcccaaccattttatttttcaatatttaccatacaaccag 2405
QY 1440 ggcctctgtcatctactataaagtctacctctgctgatccacacccatggcttctctcaacc 1499
Db 2406 ggcctctgtgtcatctactataaagtctacctctgctgatccacacccatggcttctctcaacc 2465
QY 1500 atcacctttaccagttaccttaacgacatctacatctgcgccacttctctgtctcaaaaac 1559
Db 2466 atcacctttaccagttaccttaacgacatctacatctgcgccacttctctgtctcaaaaac 2525
QY 1560 aaccttgccacagtatctcttacccttgcctgcacaggggcaagggagactcctgtgac 1619
Db 2526 aaccttgccacagtatctcttacccttgcctgcacaggggcaagggagactcctgtgac 2585
QY 1620 ttcatttcagacatcaacaatgacatcatttcatgacaaatgctccaatagtagtcaaaactgc 1679
Db 2586 ttcatttcagacatcaacaatgacatcatttcatgacaaatgctccaatagtagtcaaaactgc 2645
QY 1680 agaccttaagagccagagcccccacacccaagagaagtcatcacagaaaccaagttcaacc 1739
Db 2646 agaccttaagagccagagcccccacacccaagagaagtcatcacagaaaccaagttcaacc 2705
QY 1740 aagcctgggtgtctctgcccacagagtccacccaagctgttaacaaacaaactctcctttgcc 1799

Db 2706 aagcctgggtgtctctgcccacagagtcacccaagctgttaacaaacaaactctcctttgcc 2765
QY 1800 tccatctctaacagagtcctctccacagagcaaaccttccagccaagcaccacttagc 1859
Db 2766 tccatctctaacagagtcctctccacagagcaaaccttccagccaagcaccacttagc 2825
QY 1860 acaaatctctccaaacttccaaactacattctgaagacctctcagcctcttattgaccac 1919
Db 2826 acaaatctctccaaacttccaaactacattctgaagacctctcagcctcttattgaccac 2885
QY 1920 tcttggcacctgttcaagcacagcatctctgttgcacttggccttatagccgttacagactac 1979
Db 2886 tcttggcacctgttcaagcacagcatctctgttgcacttggccttatagccgttacagactac 2945
QY 1980 agctggaaaacagctctcgtctgacccaactcctgaaataactagttcctcaaatctcaacaga 2039
Db 2946 agctggaaaacagctctcgtctgacccaactcctgaaataactagttcctcaaatctcaacaga 3005
QY 2040 aggtggcatcagcacagaaaagaaacccagtgattgttgatgtctaccactggattgatccc 2099
Db 3006 aggtggcatcagcacagaaaagaaacccagtgattgttgatgtctaccactggattgatccc 3065
QY 2100 tttagccagtgtaaccacatacagcaaaaagaaatgaccacaagaacttggcgtttacagcaga 2159
Db 3066 tttagccagtgtaaccacatacagcaaaaagaaatgaccacaagaacttggcgtttacagcaga 3125
QY 2160 gtacagcccagcttcaagttccctcgggaacatactccttctctcccaaccacagttgtttc 2219
Db 3126 gtacagcccagcttcaagttccctcgggaacatactccttctctcccaaccacagttgtttc 3185
QY 2220 cagggctgaagacttggctcccaaatctgccaccttctgttctcagagcagcacacagtc 2279
Db 3186 cagggctgaagacttggctcccaaatctgccaccttctgttctcagagcagcacacagtc 3245
QY 2280 accaacaacactgtcctcttcagcctcagtcacacagctgtgtctgaaccttcttcttca 2339
Db 3246 accaacaacactgtcctcttcagcctcagtcacacagctgtgtctgaaccttcttcttca 3305
QY 2340 caatggcgaaatggctgcagacaacacccgctgtgtaccactgcagggtgccgccttc 2399
Db 3306 caatggcgaaatggctgcagacaacacccgctgtgtaccactgcagggtgccgccttc 3365
QY 2400 ctggcaagggagtgattgcagtggtggatggaatgagtgccctgcgaacccctgcccatc 2459
Db 3366 ctggcaagggagtgattgcagtggtggatggaatgagtgccctgcgaacccctgcccatc 3425
QY 2460 cacagccactgcacaataactcactcagggatccttctatctgaaaaatgcocgggtgggtacca 2519
Db 3426 cacagccactgcacaataactcactcagggatccttctatctgaaaaatgcocgggtgggtacca 3485
QY 2520 gtggaaaaaaggagatgtcaatttgggttagaaccttctgacagagtttataattaaagag 2579
Db 3486 gtggaaaaaaggagatgtcaatttgggttagaaccttctgacagagtttataattaaagag 3545
QY 2580 aactttcttctaaatacaactgtggaaaaacattccagactcaagaagtgtgaaaaatgagat 2639
Db 3546 aactttcttctaaatacaactgtggaaaaacattccagactcaagaagtgtgaaaaatgagat 3605
QY 2640 cacaaaaacttaaaatattgttttttcagcgtttacctagttacatccgatctacagttca 2699
Db 3606 cacaaaaacttaaaatattgttttttcagcgtttacctagttacatccgatctacagttca 3665
QY 2700 cgcctctaggaggtccaacgggtgtgtactcactcaacaacacctttccctcggcctc 2759
Db 3666 cgcctctaggaggtccaacgggtgtgtactcactcaacaacacctttccctcggcctc 3725
QY 2760 caatgtgacgctatttgacctggctgataagatgcagaaaatgtgtcaactcctcgaagtc 2819
Db 3726 caatgtgacgctatttgacctggctgataagatgcagaaaatgtgtcaactcctcgaagtc 3785
QY 2820 ctctgtcaggggtctgcagctcttctgggatctcagagggcagatcttttagagcggcagctt 2879

QY 3599 cgctgagtggaacagaggaagacagcgcatgcggggcgtagaccacagtggaagagac 3658
DB 901 cgctgagtggaacaggaaggaagacagcgcatgcggggcgtagaccacagtggaagagac 960
QY 3659 aggtggatgtggaaccacagcgctgcatactcagacaccttgttgaactggaacgtgaa 3718
DB 961 aggtggatgtggaaccacagcgctgcatactcagacaccttgttgaactggaacgtgaa 1020
QY 3719 tgtgggcccagttacaagagagctctctgagtgactgcacacatggcactggccaccaggcc 3778
DB 1021 tgtgggcccagttacaagagagctctctgagtgactgcacacatggcactggccaccaggcc 1080
QY 3779 gactattagccaggcgacacacactagacttcagtgccaggagcctggtttcccttcggtt 3838
DB 1081 gactattagccaggcgacacacactagacttcagtgccaggagcctggtttcccttcggtt 1140
QY 3839 gcacttttagtaaatgggtgggaggtttcccttttggatctctgttttgagactgttcagaa 3898
DB 1141 gcacttttagtaaatgggtgggaggtttcccttttggatctctgttttgagactgttcagaa 1200
QY 3899 agaaggctctctcccgagacacttccataggaagcaatttggatgattcttgacga 3958
DB 1201 agaaggctctctcccgagacacttccataggaagcaatttggatgattcttgacga 1260
QY 3959 aaatactggctgttaattttctcccgagcgctgcgtgctaaacacacagatgagg 4018
DB 1261 aaatactggctgttaattttctcccgagcgctgcgtgctaaacacacagatgagg 1320
QY 4019 atgagcgtacacactgaagctcgaagatctgccattgaacggacagtgatttcatatgt 4078
DB 1321 atgagcgtacacactgaagctcgaagatctgccattgaacggacagtgatttcatatgt 1380
QY 4079 tctaggttgccttatgtcacagtttccaaagccagcccccacagtgaggaatgtgtgagg 4138
DB 1381 tctaggttgccttatgtcacagtttccaaagccagcccccacagtgaggaatgtgtgagg 1440
QY 4139 caccgcacacactgcaatgtgtttttaaagtgacaggtgacacatgtattttaaagatttt 4198
DB 1441 caccgcacacactgcaatgtgtttttaaagtgacaggtgacacatgtattttaaagatttt 1500
QY 4199 ttttaaaactctcttgagtgtaaatctcaacttttccaaacgaagcctgtagcaggcaaaa 4258
DB 1501 tttttaaactctcttgagtgtaaatctcaacttttccaaacgaagcctgtagcaggcaaaa 1560
QY 4259 caacttatattgttttagctgaggtctcagcagcgacgattgacaggcaggggggcaactt 4318
DB 1561 caacttatattgttttagctgaggtctcagcagcgacgattgacaggcaggggggcaactt 1620
QY 4319 ttcatccatgaggcccaagcctgggctgggactctgatcaccattgtgagggccagag 4378
DB 1621 ttcatccatgaggcccaagcctgggctgggactctgatcaccattgtgagggccagag 1679
QY 4379 gcagctgcgtatgagggagaaatgcacactgaacgcagggtttcaccactctaggaaagc 4438
DB 1680 gcagctgcgtatgagggagaaatgcacactgaacgcagggtttcaccactctaggaaagc 1739
QY 4439 agcittgttagccctgcagctgtagtggttagaggatgggctgaatagcagggttag 4498
DB 1740 agcittgttagccctgcagctgtagtggttagaggatgggctgaatagcagggttag 1799
QY 4499 atttccctgatacaacagtgctttgggaagcgtgtgtggatttccctgaggaagacaggggagc 4558
DB 1800 atttccctgatacaacagtgctttgggaagcgtgtgtggatttccctgaggaagacaggggagc 1859
QY 4559 cgagatgagccacacatagtttgcctcacgcggtactgcagcactttgttaccagaaatc 4618
DB 1860 cgagatgagccacacatagtttgcctcacgcggtactgcagcactttgttaccagaaatc 1919
QY 4619 tcatgtccaaaaaaccocatgtaaactttcaaccactcaaaagctgtttattcggctgaaga 4678
DB 1920 tcatgtccaaaaaaccocatgtaaactttcaaccactcaaaagctgtttattcggctgaaga 1979
QY 4679 aataacttttttctcaccacagtcatttgttaacttttcatatgctgtgtcgcacccctc 4738

DB 1980 aataacttttttctcaccacagtcatttgttaacttttcatatgctgtgtcgcacccctc 2039
QY 4739 cagaacactgggttatacttccagtcagtggtggagaactgaagaacttcggttggctcgag 4798
DB 2040 cagaacactgggttatacttccagtcagtggtggagaactgaagaacttcggttggctcgag 2099
QY 4799 gaactggaaggttgacacttcgggaagagttccacatcatttatttattatccctgtgat 4858
DB 2100 gaactggaaggttgacacttcgggaagagttccacatcatttatttattatccctgtgat 2159
QY 4859 gtgggtcctgccaggagagacatccagttactcgtgtctttaaattgcccactggggaactg 4918
DB 2160 gtgggtcctgccaggagagacatccagttactcgtgtctttaaattgcccactggggaactg 2219
QY 4919 tgtttattggcctctcttggggacatccctgggttttggatgaaagtgggggaatacacagagt 4978
DB 2220 tgtttattggcctctcttggggacatccctgggttttggatgaaagtgggggaatacacagagt 2279
QY 4979 aaaaagaattgtctccacactgaagcgggagtcgcccttcacatttctggaattggtgca 5038
DB 2280 aaaaagaattgtctccacactgaagcgggagtcgcccttcacatttctggaattggtgca 2339
QY 5039 gccactggggacagttctgcccccggcgatggtttgtttcttcaaggtcctctaaataaat 5098
DB 2340 gccactggggacagttctgcccccggcgatggtttgtttctcaaggtcctctaaataaat 2399
QY 5099 cccattttacataaatccttggccctgatggtttttaagcaagaactcctgtgtcccatg 5158
DB 2400 cccattttacataaatccttggccctgatggtttttaagcaagaactcctgtgtcccatg 2459
QY 5159 gtctccacactcacatcacctcgtgtgacaaagtccttagtcagggaggtgacattt 5218
DB 2460 gtctccacactcacatcacctcgtgtgacaaagtccttagtcagggaggtgacattt 2519
QY 5219 tagtagttaaattgcaacttatccatgagataaaataaaagagaactgttttatcagtg 5278
DB 2520 tagtagttaaattgcaacttatccatgagataaaataaaagagaactgttttatcagtg 2579
QY 5279 aggttaacactaaaaattcaaaagtcgcctttttgaaatcttgggctctctctctgtag 5338
DB 2580 aggttaacactaaaaattcaaaagtcgcctttttgaaatcttgggctctctctctgtag 2639
QY 5339 aaccaatggcctttgtggtcacggcctcgacatactgagaggttctgagctcctgc 5398
DB 2640 aaccaatggcctttgtggtcacggcctcgacatactgagaggttctgagctcctgc 2699
QY 5399 agctcacctgagccacacagactaggtcttctggtccttccgcagcatgctgtcaccc 5458
DB 2700 agctcacctgagccacacagactaggtcttctggtccttccgcagcatgctgtcaccc 2759
QY 5459 coagaacccgcagctgtggaagagccatgttagggaggctaattcccaaggcatacacttcc 5518
DB 2760 coagaacccgcagctgtggaagagccatgttagggaggctaattcccaaggcatacacttcc 2819
QY 5519 actgcttccagctgacgtcacagctgcacaaatcactcctctatcggagccagaagactt 5578
DB 2820 actgcttccagctgacgtcacagctgcacaaatcactcctctatcggagccagaagactt 2879
QY 5579 cagctcccaaaaatgaagtgttctcctgaaacacattcttggggaagatcccaacatcg 5638
DB 2880 cagctcccaaaaatgaagtgttctcctgaaacacattcttggggaagatcccaacatcg 2939
QY 5639 agaaaacgggtgcctgtgagttcccaacatgcttctgttccatgggtttcttccgtatgg 5698
DB 2940 agaaaacgggtgcctgtgagttcccaacatgcttctgttccatgggtttcttccgtatgg 2999
QY 5699 agtgagttaaagagtggtttattttgttttaactgagaaaaaaggaggagcaccacaaa 5758
DB 3000 agtgagttaaagagtggtttattttgttttaactgagaaaaaaggaggagcaccacaaa 3059
QY 5759 ggttgaggtcacacagtcacagtttccagagcggtttggggggtggggaagcactt 5818

Db	3060	ggttgaggtcacacagctctccacagttctccagaggagccctttg9999t9999aaggacacct	3119
Qy	5819	ccagagcatgaggtcctaaaggagacatgagtaaaacatgctctgtgaccocagtggaggaag	5878
Db	3120	ccagagcatgaggtcctaaaggagacatgagtaaaacatgctctgtgaccocagtggaggaag	3179
Qy	5879	gagagccagctgacatctctgacccgggttccctagctgcagaaagggtcccgcttagccg	5938
Db	3180	gagagccagctgacatctctgacccgggttccctagctgcagaaagggtcccgcttagccg	3239
Qy	5939	agggaaacacctgtagcagaaaggacctgtagtcacacctggcagcgcagagctctcc	5998
Db	3240	agggaaacacctgtagcagaaaggacctgtagtcacacctggcagcgcagagctctcc	3299
Qy	5999	gccagacacagtgctccatgtcagccctgcacctgggtgctgtgattcaacgtgcacag	6058
Db	3300	gccagacacagtgctccatgtcagccctgcacctgggtgctgtgattcaacgtgcacag	3359
Qy	6059	atgccacaaacctgtgcacaaatattccacagatgggggaaggtgagaggaaggagcaagt	6118
Db	3360	atgccacaaacctgtgcacaaatattccacagatgggggaaggtgagaggaaggagcaagt	3419
Qy	6119	atgtgtaactgctcaagagatgctttaaacctcca tagagaggagccggg999caggggcat	6178
Db	3420	atgtgtaactgctcaagagatgctttaaacctcca tagagaggagccggg99caggggcat	3479
Qy	6179	ctgtgtgtcccgccacacactgcagcagggaaggtgctgctgagctccctagcatcag	6238
Db	3480	ctgtgtgtcccgccacacactgcagcagggaaggtgctgctgagctccctagcatcag	3539
Qy	6239	tgtgtgtgtttaagctccagagggtcttatgtccattgtctttccctctgccccttgagc	6298
Db	3540	tgtgtgtgtttaagctccagagggtcttatgtccattgtctttccctctgccccttgagc	3599
Qy	6299	cacctcaagccctgagctctttctttaggcagatgaactgacatgctcctaccatga	6358
Db	3600	cagctcaagccctgagctctttctttaggcagatgaactgacatgctcctaccatga	3659
Qy	6359	ccaggtctctgggcaaggctccctcacagtatcctctgagagggtgggcatggaaagtggccatt	6418
Db	3660	ccaggtctctgggcaaggctccctcacagtatcctctgagagggtgggcatggaaagtggccatt	3719
Qy	6419	tctcaggtacagaaaccttcagagaggataaataagcttgccctgtagaacagagactgaa	6478
Db	3720	tctcaggtacagaaaccttcagagaggataaataagcttgccctgtagaacagagactgaa	3779
Qy	6479	acctgtgcccgtgactccccagctactctgcccactgtagcccctctgcttactgtc	6538
Db	3780	acctgtgcccgtgactccccagctactctgcccactgtagcccctctgcttactgtc	3839
Qy	6539	ctggcacacccctcaccatcctgtataccttaaaatatacaagagggtgcaagagagaagg	6598
Db	3840	ctggcacacccctcaccatcctgtataccttaaaatatacaagagggtgcaagagagaagg	3899
Qy	6599	ctttaaga taagtatttttttaaggaaaccttaattatttttaagagagtaaccaaat	6658
Db	3900	ctttaaga taagtatttttttaaggaaaccttaattatttttaagagagtaaccaaat	3959
Qy	6659	tagtgagtgaaatgc 6674	
Db	3960	taatgacgtcaaatgc 3975	

RESULT	8
AAS02424	
ID	AAS02424 standard; cDNA; 3116 BP.
XX	
AC	AAS02424;
XX	
DT	18-JUL-2001 (first entry)
XX	
DE	Human secreted protein, cDNA #30.
XX	

KW	Human; secreted protein; Immunogen; antibody; diagnosis;
KW	rheumatoid arthritis; hyperproliferative disorder; neoplasm;
KW	cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
KW	angiogenesis; Alzheimer's disease; bacterial infection; viral infection;
KW	fungal infection; corneal infection; wound healing; cell culture;
KW	epithelial cell proliferation; skin aging; transplantation;
KW	tissue regeneration; chemotaxis; food additive; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200123546-A1.
XX	
PD	05-APR-2001.
XX	
PF	26-SEP-2000; 2000WO-US26323.
XX	
XX	27-SEP-1999; 99US-0155805.
PR	
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Rosen CA, Ruben S, Komatsoulis GA;
XX	
DR	WPI: 2001-266150/27.
DR	P-PSDB; AAU01751.
XX	
XX	Nucleic acids encoding 37 human secreted polypeptides, useful for
PT	preventing, diagnosing and/or treating e.g. cancers, Parkinson's
PT	disease and diabetic retinopathy -
XX	
PS	Disclosure; Page 432-433; 494pp; English.
XX	
CC	The sequence encodes a human secreted protein of the invention. The
CC	polynucleotides, polypeptides and antibodies raised against them are used
CC	to prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC	rabbits, goats, horses, cats, dogs, chickens or sheep. The
CC	polynucleotides and antibodies are also used in diagnosing a pathological
CC	condition or susceptibility to a pathological condition. The antibodies
CC	can also be used in alleviating symptoms associated with the disorders
CC	and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC	immunosorbent assays (ELISA). Disorders which are diagnosed or treated
CC	include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC	disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC	e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC	angiogenesis, nervous system disorders e.g. Alzheimer's disease,
CC	infections caused by bacteria, viruses and fungi and ocular disorders
CC	e.g. corneal infection. The polypeptides can also be used to aid wound
CC	healing and epithelial cell proliferation, to prevent skin aging due to
CC	sunburn, to maintain organs before transplantation, for supporting cell
CC	culture of primary tissues, to regenerate tissues and in chemotaxis. The
CC	polypeptides can also be used as a food additive or preservative to
CC	increase or decrease storage capabilities.
XX	
SO	Sequence 3116 BP: 770 A: 743 C: 810 G: 793 T: 0 other;

[illegible]

Db 182 acctggcactggcaccagggcgactatttagccaggggcagaccactagacttccagtgcag 241
QY 3817 ggaacctggtttcccttcgtttgcaccttagtaataattgggtggagggtttcccttttggat 3876
Db 242 ggaacctgggttcccttcgtttgcaccttagtaataattgggtggagggtttcccttttggat 301
QY 3877 ctgttttgagactgtttccagaagaagggtccctttcccgagagacacttccatagggcagca 3936
Db 302 ctgttttgagactgtttccagaagaagggtccctttcccgagagacacttccatagggcagca 361
QY 3937 atttgggtatcatttgcagcaaaaaactgtggtctgtttaatttttccctgcccagcgct 3996
Db 362 atttgggtatcatttgcagcaaaaaactgtggtctgtttaatttttccctgcccagcgct 421
QY 3997 gctgtctaaacaacagatgaggatgagcgtaccactgaactctgaagatgtcgcatttga 4056
Db 422 gcgtctaaacaacagatgaggatgagcgtaccactgaactctgaagatgtcgcatttga 481
QY 4057 acggacagtgttttcatagtttctaggttctgtctatgtctacagtgttccaagccagcccc 4116
Db 482 acggacagtgttttcatagtttctaggttctgtctatgtctacagtgttccaagccagcccc 541
QY 4117 cacagtggagaaatgtgtgagggcaccgcacacaactgcaatgtgtttttaaagtcgaagt 4176
Db 542 cacagtggagaaatgtgtgagggcaccgcacacaactgcaatgtgtttttaaagtcgaagt 601
QY 4177 gacacatgtatttaagaattttttttaaatactctttgcaagttaaactcactttttcaa 4236
Db 602 gacacatgtatttaagaattttttttaaatactctttgcaagttaaactcactttttcaa 661
QY 4237 acaagctggatcagggcaaaaaacaacttatatttgtgttttagctgaggtctagcaggca 4296
Db 662 acaagctggatcagggcaaaaaacaacttatatttgtgttttagctgaggtctagcaggca 721
QY 4297 gattgcaggcaggggggcaacttttccatcattgagggcccgacctgggctgtggaactctg 4356
Db 722 gattgcaggcaggggggcaacttttccatcattgagggcccgacctgggctgtggaactc-g 780
QY 4357 atcacactgtgagggccagagcagctgcgtatgaggagagaataatgtcaaaactgaacgca 4416
Db 781 atcacactgtgagggccagagcagctgcgtatgaggagagaataatgtcaaaactgaacgca 840
QY 4417 ggtttcaccaacttaggaagcagctgtgtgagccctgcagctgcagtgtggttagaggg 4476
Db 841 ggtttcaccaacttaggaagcagctgtgtgagccctgcagctgcagtgtggttagaggg 900
QY 4477 atgggctgaatagggcaggttagatttccctgcatacaacagtgccttgggaagcgtgtgga 4536
Db 901 atgggctgaatagggcaggttagatttccctgcatacaacagtgccttgggaagcgtgtgga 960
QY 4537 ttccctgaggaagaacagggagccagatggagccacacatgatttgcaccggctact 4596
Db 961 ttccctgaggaagaacagggagccagatggagccacacatgatttgcaccggctact 1020
QY 4597 gcagcactttgtaccagaaatctcatgtccacaaaacccatgtaaaactttcaaccactca 4656
Db 1021 gcagcactttgtaccagaaatctcatgtccacaaaacccatgtaaaactttcaaccactca 1080
QY 4657 aagctgtttattcgtgtgaagaaataacttttttttccaccagtcatttgttacctctt 4716
Db 1081 aagctgtttattcgtgtgaagaaataacttttttttccaccagtcatttgttacctctt 1140
QY 4717 catatgctgtgtcgcaacctccagaaacgtgtttatacttccagtcagtgtaggaac 4776
Db 1141 catatgctgtgtcgcaacctccagaaacgtgtttatacttccagtcagtgtaggaac 1200
QY 4777 tgaagacttccggttggctgaggaactgaggttgcaccttcgggaagggaagtcccaactca 4836
Db 1201 tgaagacttccggttggctgaggaactgaggttgcaccttcgggaagggaagtcccaactca 1260
QY 4837 tcttatttattatgcctgtgagtgtgggtccctgcaggagagacatccagttactcgtgtct 4896
Db 1261 tcttatttattatgcctgtgagtgtgggtccctgcaggagagacatccagttactcgtgtct 1320

QY 4897 ttaattgccacctggggaaactgtgttttatggcccttctttggggcaccctcctggttttggat 4956
Db 1321 ttaattgccacctggggaaactgtgttttatggcccttctttggggcaccctcctggttttggat 1380
QY 4957 gaagtggagggaaacacagaggtaaaaaattgtctccaccctgaagcgggagtcctcgct 5016
Db 1381 gaagtggagggaaacacagaggtaaaaaattgtctccaccctgaagcgggagtcctcgct 1440
QY 5017 tcaatttctggaaaatgggtgcagccactggggacagtctctgccccgggacatggttgtttc 5076
Db 1441 tcaatttctggaaaatgggtgcagccactggggacagtctctgccccgggacatggttgtttc 1500
QY 5077 tcaaggtctctctaaataataatccctattcttacataatcccttggccctgatgttttaa 5136
Db 1501 tcaaggtctctctaaataataatccctattcttacataatcccttggccctgatgttttaa 1560
QY 5137 gcaagaactcctgtgtccccatggtctccacactccacatccacactcctctgttagcaagat 5196
Db 1561 gcaagaactcctgtgtcccatggtctccacactccacatccacactcctctgttagcaagat 1620
QY 5197 cctagtcaggggaggtgcatttttagtagttaaatgtcaacttatccatgagataataaaa 5256
Db 1621 cctagtcaggggaggtgcatttttagtagttaacattgcacttatccatgagataataaaa 1680
QY 5257 ggagaactgttttatacagtggaggttaaccttaaaaatttcaaaagtctcgcccttttga 5316
Db 1681 ggagaactgttttatacagtggaggttaaccttaaaaatttcaaaagtctcgcccttttga 1740
QY 5317 tcttggccctctctctctgtagaaacaaatggcccttctgtggtcacagcctcgcacactaa 5376
Db 1741 tcttggccctctctctctgtagaaacaaatggcccttctgtggtcacagcctcgcacactaa 1800
QY 5377 ctggagagtctgagctcctgcagctcacctgagccacagactaggtcttcttggctcct 5436
Db 1801 ctggagagtctgagctcctgcagctcacctgagccacagactaggtcttcttggctcct 1860
QY 5437 tccgcagcatgctgtctacccccagaaacccgcagctgtgggaagagccatgtaggaggg 5496
Db 1861 tccgcagcatgctgtctacccccagaaacccgcagctgtgggaagagccatgtaggaggg 1920
QY 5497 ctattccaggtacatacttccactgccttcagctgcagctgcagctgcacagctgacaaaatcatctc 5556
Db 1921 ctattccaggtacatacttccactgccttcagctgcagctgcagctgcacagctgacaaaatcatctc 1980
QY 5557 ctctatcgggccagaagacttccagctccacaaaatgaagtgttctctcctgaaaaacatt 5616
Db 1981 ctctatcgggccagaagacttccagctccacaaaatgaagtgttctctcctgaaaaacatt 2040
QY 5617 ctgggaagaatcccaacatcgagaacacgtgtccctgtgagttcccaacatgctcttg 5676
Db 2041 ctgggaagaatcccaacatcgagaacacgtgtccctgtgagttcccaacatgctcttg 2100
QY 5677 ttcatgggtttcttcgtatggagtggattaaagagtgttttattttgttcttaactga 5736
Db 2101 ttcatgggtttcttcgtatggagtggattaaagagtgttttattttgttcttaactga 2160
QY 5737 gaaaaaaggagggcaccacaaagttgaggttcacacagttccacagtttccagggagggc 5796
Db 2161 gaaaaaaggagggcaccacaaagttgaggttcacacagttcctcagtttccagggagggc 2220
QY 5797 ttgggggtgggaaggcaccctccagagcatgaggtcttaaggggacatgagtaaacat 5856
Db 2221 ttgggggtgggaaggcaccctccagagcatgaggtcttaaggggacatgagtaaacat 2280
QY 5857 gtcgtgtaccocaggtgaggaaggagagccagctgcactcctgcacggggttccctagctg 5916
Db 2281 gtcgtgtaccocaggtgaggaaggagagccagctgcactcctgcacggggttccctagctg 2340
QY 5917 cagaaagggtccccgcctaggccgaggggaaaaacactgtagcagaagaggcctggatgcac 5976
Db 2341 cagaaagggtccccgcctaggccgaggggaaaaacactgtagcagaagaggcctggatgcac 2400

QY 4057 acggacagtggttttcataatgtttctagagttgtcttattgtctacagtttccaaagccagcccc 4116
Db 524 acggacagtggttttcataatgtttctagagttgtcttattgtctacagtttccaaagccagcccc 583
QY 4117 cacagtgaggaaaatgtgtgagggcacccacacaaact-gcaatgtgttttttaagtcaagg 4175
Db 584 cacagtgaggaaaatgtgtgagggcacccacacaaactggcaaatgtgttttttaagtcacaagg 643
QY 4176 tgacacatgtatttaaga-ttttttttaaaatctctttgcagttaaatctcaactttttc 4234
Db 644 tgacacatgtatttaagaattttttttaaaatctctttgcagttaaatctcaactttttc 703
QY 4235 aaacaagcctggatcagggcgaacaaacttatattgttttttagctgagagctcagcagg 4294
Db 704 aaacaagcctggatcagggcgaacaaacttatattgttttttagctgagagctcagcagg 763
QY 4295 cagattcagcgaggggggcactttttcatccatgagggccagcctggggcctgggactc 4354
Db 764 cagattcagcgaggggggcactttttcatccatgargggccagcctggggcctgggactc 823
QY 4355 tgatcacattgtgagggccagaggcgagctgcgtatgtaggagaaaatgtcaaaactgaacg 4414
Db 824 tgatcacattgtgagggccagaggcgagctgcgtatgtaggagaaaatgtcaaaactgaacg 883
QY 4415 cagggttcacacactctaggaagcagctgtgtgagccctgcagctgagatgtgttagag 4474
Db 884 cagggttcacacactctaggaagcagctgtgtgagccctgcagctgagatgtgttagag 943
QY 4475 ggatgggctgaataggcaggttagatttctctgcatcaacagtgctttgggaagctggtg 4534
Db 944 ggatgggctgaataggcaggttagatttctctgcatcaacagtgctttgggaagctggtg 1003
QY 4535 gattcctgaggaagaaacagggagccggagatggagccacacatagtttgcaccggcta 4594
Db 1004 gattcctgaggaagaaacagggagccggagatggagccacacatagtttgcaccggcta 1063
QY 4595 ctgcagcactttgtaccagaaatctcatgtccacaaacccatgtaaacttcaacact 4654
Db 1064 ctgcagcactttgtaccagaaatctcatgtccacaaacccatgtaaacttcaacact 1123
QY 4655 caaagctgtttattcggtgaaagaaaataactttttttctcaccagctcatttgcactc 4714
Db 1124 caaagctgtttattcggtgaaagaaaataactttttttctcaccagctcatttgcactc 1183
QY 4715 ttcatatgctgtgtctgcaccctccagaaaacgttggttatacttccagtcagttgggaga 4774
Db 1184 ttcatatggtgtatgtctgcaccctccagaaaacgttggttatacttccagtcagttgggaga 1243
QY 4775 actgaagacttcgggttcgaggaaactgagggttgaccttcgggaaggaaagtccact 4834
Db 1244 actgaagacttcgggttcgaggaaactgagggttgaccttcgggaaggaaagtccact 1303
QY 4835 catcttatttattatgctgtg-atgtgggtctctgcaggaggaacatccagtaactcgttg 4893
Db 1304 catcttatttattatgctgtgaatgtgggtctctgcaggaggaacatccagtaactcgttg 1363
QY 4894 tctttaattgccacctggggaactgtgtttattggccctttttggggcatcctgggtttg 4953
Db 1364 tctttaattgccacctggggaactgtgtttattggccctttttggggcatcctgggtttg 1423
QY 4954 gatgaagtgggggaaatcacagaggtaaaagaattgtctccacccctgaagcgggagtcgcc 5013
Db 1424 gatgaagtgggggaaatcacagaggtaaaagaattgtctccacccctgaagcgggagtcgcc 1483
QY 5014 gcttcacatttctggaaatggtgcagccactggggagacttctgccccgggcatggttgt 5073
Db 1484 gcttcacatttctggaaatggtgcagccactggggagacttctgccccgggcatggttgt 1543
QY 5074 ttcttcaaggtctcttaaaatataatccctattcttacaataatccttggccctgatggttt 5133
Db 1544 ttcttcaaggtctcttaaaatataatccctattcttacaataatccttggccctgatggttt 1603

QY 5134 taagcaagaactcctgtgtgtcccatggtctccacacactcaccatcaccctgtgttagcaag 5193
Db 1604 taagcaagaactcctgtgtgtcccatggtctccacacactcaccatcaccctgtgttagcaag 1663
QY 5194 agtctcagtcaggggaggtgcatttttagtagttaaatgtcaattatccatgagataaata 5253
Db 1664 agtctcagtcaggggaggtgcatttttagtagttaamatgtcaattatccatgagataaata 1723
QY 5254 aaagggaactgtttttatcagtgggaggtcaacctaaatttcaaaatgtcgcctttttg 5313
Db 1724 aaagggaactgtttttatcagtgaggctaaactaaatttcaaaatgtcgcctttttg 1783
QY 5314 aaatcttgggacctctctctctgtagaaccaatggccctttgttggtcaagcgcctcgacc 5373
Db 1784 aaatcttgggacctctctctctgtagaaccaatggccctttgttggtcaagcgcctcgacc 1843
QY 5374 taactggagagttctgagctcctcagctcactgcagcccacagactaggctttctggct 5433
Db 1844 taactggagagttctgagctcctcagctcactgcagcccacagactaggctttctggct 1903
QY 5434 ccttcgcagcatgctctgtctcacccccagaaacccgcagctgtgggaagagccaatgtagg 5493
Db 1904 ccttcgcagcatgctctgtctcacccccagaaacccgcagctgtgggaagagccaatgtagg 1963
QY 5494 aggcatttccaggcatatacacttccactccttcagctgcagctgcacagctgacaaaatcat 5553
Db 1964 aggcatttccaggcatatacacttccactccttcagctgcagctgcacagctgacaaaatcat 2023
QY 5554 ctctctatcggagccagagaagacttcagctccacaaatgaagtgtctctgcctgaaaaac 5613
Db 2024 ctctctatcggagccagagaagacttcagctccacaaatgaagtgtctctgcctgaaaaac 2083
QY 5614 attcttgggaagaaatcccaaatcagaaaaacggtgtctctgtgagttccaaacaatgcttc 5673
Db 2084 attcttgggaagaaatcccaaatcagaaaaacggtgtctctgtgagttccaaacaatgcttc 2143
QY 5674 ttgtctatggtttcttccgtatgagtgatttaagagtttttattttgtttcttaac 5733
Db 2144 ttgtctatggtttcttccgtatgagtgatttaagagtttttattttgtttcttaac 2203
QY 5734 tgagaaaaaaggagggcaccccaaaagttgaggttcacacagctctccacagtttccagag 5793
Db 2204 tgagaaaaaaggagggcaccccaaaagttgaggttcacacagctctccacagtttccagag 2263
QY 5794 gcgtttgggggtgggggaagcaccctccagagcatgaggctctaaagggagacatgagtaaa 5853
Db 2264 gcgtttgggggtgggggaagcaccctccagagcatgaggctctaaagggagacatgagtaaa 2323
QY 5854 catgtctgtgaccagtgaggaaaggagagccagctgcactcctgcacgggttccctag 5913
Db 2324 catgtctgtgaccagtgaggaaaggagagccagctgcactcctgcacgggttccctag 2383
QY 5914 ctgcagaaggggtccgccttagggccgaggggaaacacactgatagcagaagagcctgagtg 5973
Db 2384 ctgcagaaggggtccgccttagggccgaggggaaacacactgatagcagaagagcctgagtg 2443
QY 5974 cacacctgcacgcggaggtctctccgcccagagacacagtgctcctatgtcagccccctgacc 6033
Db 2444 cacacctgcacgcggaggtctctccgcccagagacacagtgctcctatgtcagccccctgacc 2503
QY 6034 tgggggtgtgtattccagtgacagatgccacaatcctgcaccaatctccacacagatggg 6093
Db 2504 tgggggtgtgtattccagtgacagatgccacaatcctgcaccaatctccacacagatggg 2563
QY 6094 ggaaggtgagagaaggggcaagtgtgttaactgctcaagagatgcttaaacctccat 6153
Db 2564 ggaaggtgagagaaggggcaagtgtgtgttaactgctcaagagatgcttaaacctccat 2623
QY 6154 agagagagcggggcagggggaatctgtgtgtcccgctcacacactgcagcagggaaggg 6213
Db 2624 agagagagcggggcagggggaatctgtgtgtcccgctcacacactgcagcagggaaggg 2683
QY 6214 tggctgggtggctccctcctggcatcagtggtttgttttaagctccagaggggtcttattgcca 6273

Db 2684 tggctggctggcccttgccatcagtggtttgtttaaagctccagagggtctctattgcca 2743
QY 6274 ttgtcttttctctgccccttgagccagcctaagggcccttgagctctgtttctttaggcgg 6333
Db 2744 ttgtcttttctctgccccttgagccagcctaagggcccttgagctctgtttctttaggcgg 2803
QY 6334 atgaactgacatgctcctaccatgaccaggtctctggcagggtctcctcagtatccttg 6393
Db 2804 atgaactgacatgctcctaccatgaccaggtctctggcagggtctcctcagtatccttg 2863
QY 6394 agagggtggcatagaagtgcccatttctcaggtacagaaaaccttcagagaggataatag 6453
Db 2864 agagggtggcatagaagtgcccatttctcaggtacagaaaaccttcagagaggataatag 2923
QY 6454 ctgtccctgtagaagcaggagactgaacacctgttccgctgactcccccagctactctgcc 6513
Db 2924 ctgtccctgtagaagcaggagactgaacacctgttccgctgactcccccagctactctgcc 2983
QY 6514 cactgtagccctgcttactgtctctggtcacacacctcaccatcctgtataccttaaat 6573
Db 2984 cactgtagccctgcttactgtctctggtcacacacctcaccatcctgtataccttaaat 3043
QY 6574 atcaagagggaagagagaaagggtttaagataaagttattttttaaggaaaccttaa 6633
Db 3044 atcaagagggaagagagaaagggtttaagataaagttattttttaaggaaaccttaa 3103
QY 6634 tattatttttaagaagtaaaccaattagtgacgtgaaatgcataaaataaaataaaat 6693
Db 3104 tattatttttaagaagtaaaccaattagtgacgtgaaatgcataaaataaaataaaatn 3163
QY 6694 gctgact 6700
Db 3164 gctgact 3170

RESULT 10
AAL26437
ID AAL26437 standard; cDNA; 3075 BP.
XX
AC AAL26437;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 18894.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US00798.
XX
PR 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS Claim 1; Page 3478-3479; 3695pp; English.

XX The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterising treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
XX activity.
SQ Sequence 3075 BP; 844 A; 653 C; 717 G; 854 T; 7 other;

Query Match 12.0%; Score 831; DB 22; Length 3075;
Best Local Similarity 89.1%; Pred. No. 2.8e-217;
Matches 997; Conservative 0; Mismatches 90; Indels 32; Gaps 8;
QY 5828 gaggctctaaggggacatgagtaaacatgtctgtgacccagtgagaaaggagagacca 5887
Db 2 gaggctctaaggggacatgagtaaacatgtctgtgacccagtgagaaaggagagacca 61
QY 5888 gctgcactctgcacaggggttctctagctgcagaagg--tcccgcctagccgagggga 5944
Db 62 gctgcactctgcaggggttctctagctgcagaagggttcccgcgcagagcccgggggg 121
QY 5945 aaca---ccctgtagcagaagagggcctggatgcacacctggcacgcgcgaggtctccgc 6000
Db 122 gacaacacctgaaagcagaagggccctggaaatgaacacctggcacgcgcgagggctc 181
QY 6001 ccagacacagtgctccatgtcagccctg-----cacctgggggtgtgtgat 6046
Db 182 cccgcgcccaagacacaggtctgcacctgtcaaccccttgcaacctgggtttgtgat 241
QY 6047 tcaagtg---cacagatgccaaa---tctctgaccaatatccacagatgggggaag 6098
Db 242 ttcacgttgcaaaaagatgcaccaatcctgcgcccccacaaatcccaacagatgggggaag 301
QY 6099 gtgagagggaaggggcaagtgtgttaactgtctcaagagatgtcttaacctccatagaga 6158
Db 302 gtgagagggaagggcgaagtgtgttaactgtctcaagagatgtcttaacctccatagaga 361
QY 6159 ggagccggggcgagggggcatctgtgtcccgctcaacacctgcagcagggaaagggtgct 6218
Db 362 ggagccggggcgagggggcatctgtgtcccgctcaacacctgcagcagggaaagggtgct 421
QY 6219 gctgtgctccctgcatcagtggtttggttaagctccagaggtcttatgtccattgtc 6278
Db 422 ggtgtgctccctgcatcagtggtttggttaagctccagaggtcttatgtccattgtc 481
QY 6279 ttttctctgccccttgagccagcctaagggccctggagctgtttcttaggcggatgaa 6338
Db 482 ttttctctgccccttgagccagcctaagggccctggagctgtttcttaggcggatgaa 541
QY 6339 ctgacatgctctaccatgaccaggtctctggcaggctcctcacagatcctctgagagg 6398
Db 542 ctgacatgctctaccatgaccaggtctctggcaggctcctcacagatcctctgagagg 601
QY 6399 tgggcatggaagtgccattttcaggtacagaaaaccttcagagagataaataagcttc 6458
Db 602 tgggcatggaagtgccattttcaggtacagaaaaccttcagagagataaataagcttc 661
QY 6459 cctgtagaagcaggactgaaacctgtccgctgactccccccagctactctgccactg 6518
Db 662 cctgtagaagcaggactgaaacctgtccgctgactccccccagctactctgccactg 721
QY 6519 tagccccctgcttactgtcctggcacacccctccacatcctgtatataccttaataatcaa 6578
Db 722 tagccccctgcttactgtcctggcacacccctccacatcctgtatataccttaataatcaa 781
QY 6579 agagggcagaagagaaaggcctttaagataaagttattttttaaggaaaccttaataat 6638
Db 782 agagggcagaagagaaaggcctttaagataaagttattttttaaggaaaccttaataat 841

XX 05-SEP-2001.
PD 07-JUL-2000; 2000BP-0114089.
PF 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX (HELI-) HELIX RES INST.
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2001-524255/58.
XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
PT Example 11; SEQ ID NO 2163; 1380pp + sequence listing; English.
PS The invention relates to primers for synthesising full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence was used as the
CC representative sequence from a human clone which was used in
CC homology searches to identify the clone.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX Sequence 751 BP; 198 A; 187 C; 188 G; 174 T; 4 other;
SQ

Query Match 10.2%; Score 708.2; DB 22; Length 751;
Best Local Similarity 98.7%; Pred. No. 6.7e-184;
Matches 743; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

QY 2699 acgcctctaggagtcacacgcggttgatctcaactgcaaacacacctttccctggcct 2758
Db 1 acgcctctaggagtcacacgcggttgatctcaactgcaaacacacctttccctggcct 60
QY 2759 ccaatgtgacgctatttgacctgctgagatgagatgcagaaatgttcaactcctgcaagt 2818
Db 61 ccaatgtgacgctatttgacctgctgagatgagatgcagaaatgttcaactcctgcaagt 120
QY 2819 cctctgctgaggtctgcagctcttggatctcagagcgcgatcttttagagcgggcagct 2878
Db 121 cctctgctgaggtctgcagctcttggatctcagagcgcgatcttttagagcgggcagct 180
QY 2879 tgtgcaagcggaagatcccgatgtgacaaagacacccctccatctcactgacactggagc 2938
Db 181 tgtgcaagcggaagatcccgatgtgacaaagacacccctccatctcactgacactggagc 240
QY 2939 gcgttgccctgtgccagtgcaagtcggtgatacttccagttcaacagatggaccactcct 2998
Db 241 gcgttgccctgtgccagtgcaagtcggtgatacttccagttcaacagatggaccactcct 300
QY 2999 gccgagcatgtgaagatggatagcttgaaatgaaacctgcagtgagtgcccatcttg 3058
Db 301 gccgagcatgtgaagatggatagcttgaaatgaaacctgcagtgagtgcccatcttg 360
QY 3059 gccttggtggtctcaactgtggaacccctatcagcttatcactgtggtgatcgagcgcc 3118
Db 361 gccttggtggtctcaactgtggaacccctatcagcttatcactgtggtgatcgagcgcc 420
QY 3119 cggaggtgggctcctgctcactatcctaggatcgacactgattgttactctgttcagaaaga 3178
Db 421 cggaggtgggctcctgctcactatcctaggatcgacactgattgttactctgttcagaaaga 480

QY 3179 ataaaaatgacataagcaaaactcatcttcaaaaagtgagagatttccaaaatgtcccccatg 3238
Db 481 ataaaaatgacataagcaaaactcatcttcaaaaagtgagagatttccaaaatgtcccccatg 540
QY 3239 ctgaatacccccacaaaatcctcgtctcaacaagaatgggcccagagaagctattgaaatgcag 3298
Db 541 ctgaatacccccacaaaatcctcgtctcaacaagaatgggcccagagaagctattgaaatgcag 600
QY 3299 agaattggaagtaccacaaaacacctctcctccagatgacggatgtgtactactcctcctacaagt 3358
Db 601 agaattggaagtaccacaaaacacct-ctncagatgacggatgtgtactactcg-ctacaagt 658
QY 3359 taaggaaatccagaacttgaacgaacggac-tctaccggcgctcacactggaactgcccagga 3417
Db 659 taaggaaatccagaacttgaacgaacggacttctaccggcgnctcacactggaactgcccagga 718
QY 3418 tcacggcattcttgcatcttcccccggacagat 3450
Db 719 tcacgggattcttgcatcttcccccggacagat 751

RESULT 13
AAH31088
ID AAH31088 standard; cDNA; 869 BP.
XX AAH31088;
AC AAH31088;
XX 27-JUL-2001 (first entry)
XX Human colon cancer cell line Km12L4-A cDNA library derived seq#1022.
DE Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;
XX KW detection; colon cancer cell line Km12L4-A; ss.
XX OS Homo sapiens.
XX WO200018916-A2.
XX PD 06-APR-2000.
XX PF 23-SEP-1999; 99WO-US222226.
XX PR 28-SEP-1998; 98US-01021161.
PR 28-SEP-1998; 98US-01021180.
PR 29-SEP-1998; 98US-0102380.
PR 08-OCT-1998; 98US-0103815.
PR 27-OCT-1998; 98US-0105877.
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSEQ INC.
XX Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI Reinhard C, Glese K, Randazzo F, Kennedy GC, Pot D, Kassem A;
PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX WPI; 2000-293155/25.
XX Polynucleotide library comprising 1079 defined sequences, useful in
PT the form of an array to detect cancer or susceptibility to cancer -
XX Claim 1; Page 479-480; 502pp; English.
XX The present invention describes a library of polynucleotides comprising
CC 1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described
CC are: (1) an isolated polynucleotide (I) having at least 90% identity to
CC one of the 1079 sequences; (2) a recombinant host cell containing (1);
CC (3) an isolated polypeptide (II) encoded by (1); (4) an antibody that
CC specifically binds to (II); (5) a vector comprising (1); and (6) a method
CC of detecting differentially expressed genes correlated with a cancerous
CC state of a mammalian cell comprising detecting a gene product encoded by
CC 65 of the 1079 sequences given in the specification. The polynucleotides

CC are used to monitor patients having (or susceptible) to cancer to detect
CC potentially malignant events at a molecular level before they are
CC detectable at a gross morphological level. The polynucleotides are also
CC useful for monitoring the efficacy of various therapies and preventive
CC interventions. Polynucleotide probes based on the disclosed sequences
CC are useful for chromosome mapping and detection of transcription levels.
CC The 1079 polynucleotide sequences were derived from a human colon cancer
CC cell line Kml12L4-A cDNA library.
XX
SQ Sequence 869 BP; 221 A; 224 C; 235 G; 185 T; 4 other;

Query Match 10.1%; Score 703; DB 21; Length 869;
Best Local Similarity 94.2%; Pred. No. 2e-182;
Matches 826; Conservative 2; Mismatches 34; Indels 15; Gaps 9;
QY 5813 gcacctcagagcatgagctcctaagggacatgagtaagcatgtctgtgaccacagtga 5872
Db 1 gcacctcagagcatgagctcctaagggacatgagtaagcatgtctgtgaccacagtga 60
QY 5873 ggaaggagagggcagctgcactctgcacggggttctctagctgacaaaggtcccgct 5932
Db 61 ggaaggagagggcagctgcactctgcacggggttctctagctgacaaaggtcccgct 120
QY 5933 agggcaggggaaacacctgtagcagaagagggcctggatgcacacctggcagccgagg 5992
Db 121 agggcaggggaaacacctgtagcagaagagggcctggatgcacacctggcagccgagg 180
QY 5993 ctctcccccagacacagtgctccatgtcagccctgcacctggggtgtgtgattcacgt 6052
Db 181 ctctcccccagacacagtgctccatgtcagccctgcacctggggtgtgtgattcacgt 240
QY 6053 gcacagatgcacaaatcctgcacaaatattccacagatggggaa-ggtgagagaaggg 6111
Db 241 gcacagatgcacaaatcctgcacaaatattccacagatggggaa-ggtgagagaaggg 300
QY 6112 gcaagtgtgttaactgtcctaagagatgcttaaacctccatagagagccggcgca 6171
Db 301 gsaagtgtgttaactgtcctaagagatgcttaaacctccatagagagccggcgca 360
QY 6172 ggggcatctgtgtcccgctcacacatgcagcaggggaaaggggtggctggctccctg 6231
Db 361 ggggcatctgtgtcccgctcacacatgcagcaggggaaaggggtggctggctccctg 420
QY 6232 qcacagtggtgtgttaagctccagaggtcttattgcattgtctttctctctccc 6291
Db 421 qcacagtggtgtgttaagctccagaggtcttattgcattgtctttctctctccc 480
QY 6292 cttgagcagcctaagggcctggagtctgtttctttt-aggcggatgaactgacatgctcc 6350
Db 481 cttgagcagcctaagggcctggagtctgtttctttt-aggcggatgaactgacatgctcc 540
QY 6351 tacc- --tgaccaggtctctgggc- aaggtctctcacagatccttgag- aggtgggcatg 6406
Db 541 tacc- --tgaccaggtctctgggc- aaggtctctcacagatccttgag- aggtgggcatg 600
QY 6407 gaagtgcacattctcaggtacagaaacaccttcagagaggataaatagc- ttgcctcttag 6465
Db 601 gaagtgcacattctcaggtacagaaacaccttcagagaggataaatagc- ttgcctcttag 660
QY 6466 aagcaggactgaaacctgtgcccctgactccccagctactctgcccactgtagcccc 6525
Db 661 aagcaggactgaaacctgtgcccctgactccccagctactctgcccactgtagcccc 720
QY 6526 ctgccttactgtctggcaccacctcaccatctctgtatacccttaataatacaagagggc 6585
Db 721 ctgccttactgtctggcaccacctcaccatctctgtatacccttaataatacaagagggc 780
QY 6586 aagagagaagggccttaagataagttatttttttaaggaaccttaattatttttaa 6645
Db 781 aagagagaaggg-cttaagataagttatttttt- --aaggaccttaataatttttaa 834
QY 6646 gaagtaaccaaatagtgagcgtgaaatgcaaaaaa 6682

Db 835 gaagtaccacaa--tagtggcgtgnaatgccccaaaaa 869
RESULT 14
AAL34960/c
ID AAL34960 standard; cDNA; 537 BP.
XX
AC AAL34960;
XX
DT 08-JAN-2002 (first entry)
XX Human musculoskeletal system related polynucleotide SEQ ID NO 302.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antinflammatory; antitumor;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ss.
XX
OS Homo sapiens.
XX
PN WO200155367-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01338.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.

PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 05-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 P-PSDB; ABB03378.
 DR WPI; 2001-451937/48.
 DR P-PSDB; ABB03378.
 XX
 Isolated polypeptide for treating, preventing and/ or prognosing
 PT disorders related to the musculoskeletal system including
 PT musculoskeletal cancers and also for testing and detection e.g.
 PT diagnosis -
 PT
 XX
 Claim 1: SEQ ID NO 302; 781pp + Sequence Listing; English.
 XX
 The invention relates to novel genes (AAL34669-AAL37666) and proteins
 CC (ABB03087-ABB04109) associated with the musculoskeletal system useful
 CC for preventing, treating or ameliorating medical conditions e.g. by
 CC protein or gene therapy. The genes are isolated from a range of human
 CC tissues disclosed in the specification. The nucleic acids, proteins,
 CC antibodies and (ant)agonists are useful in the diagnosis, treatment
 CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
 CC other cancers of the adrenal gland, bone, bone marrow, breast,
 CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
 CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
 CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
 CC and (f) infectious diseases such as viral, bacterial, fungal and
 CC parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pt_sequences.
 XX
 SQ Sequence 537 BP; 137 A; 137 C; 134 G; 123 T; 6 other;

Query Match 7.2%; Score 500.8; DB 22; Length 537;
 Best Local Similarity 97.9%; Pred. No. 5.8e-127;
 Matches 524; Conservative 1; Mismatches 8; Indels 2; Gaps 2;

QY 2633 atgagatcaccaaaacgttaaatgtgttttcacggttacctagttacatccgatcta 2692
 |||||
 Db 537 ATGAGATCACCAAAACGTTAAATATGTGTTTTCAGCGTTACNTAGTAACATCCGATCTA 478
 QY 2693 cagttcacgccttagggaggtcccaacgcggtgtgtatctcactgcgaacaacctttccc 2752
 |||||
 Db 477 CAGTTCACGCTTCTAGGGAGTCCCAACGCGGTGGTGTATCTCACTGCAAAACNACNTTTTCCC 418
 QY 2753 tggcctccaatgtgacgcctatttgacctgacgtgctgtagtaggcagaaagtgtcaactcct 2812
 |||||
 Db 417 TGGCCTCCAATGACGCGCTATTTGACCTGGCTGATAGGATGCAGAAATGTGTCAACTCCT 358

```

QY 2813 gcaagtc-ctctgctgaggtgtccagct-cttggggtctcagagcgcgatcttttagc 2870
Db 357 GCAAGTCTTTCTGCTGAGGCTNCCACGCTCTTGGGATCTCAGAGCGGATCTTTAGAGC 298

QY 2871 gggcagcttgcgaagcggaagtcgccgaatgtgacaaagacacacctccatctgcactga 2930
Db 297 GGCAGAGCTTGTGCAAGCGGAAGAGTCCCGAATGTGACAAAGACACACCTCCATMTGCACTGA 238

QY 2931 cctggacggcgttgcctgtgcgaagtcaagtcgaggtcgggatactttcagttcaacaagatgga 2990
Db 237 CCGTGGAGCGGCTTGGCCCTGTGCCAGTGTCAAGTCCGGGATCTTTTCAGTTCACCAAGATGGA 178

QY 2991 ccactctcgcgagcatgtgaagatgataggtcttgaataatgaaacctgcactgagttg 3050
Db 177 CCACCTCTGCGGACATGTGAAGTGTATAGGCTTGAAATGAAACCTTGCACTGATGATG 118

QY 3051 cccatttggccttgggtgtcactgtggaacccctatcagcttatcactgtggtgat 3110
Db 117 CCCATTGGGCTTGGTGTCTCAACTGTGGAACCCCTATCAGCTTATCAGTGTGTTGAT 58

QY 3111 cgaagccgggaggtgggtcctctgtcctccttaggcatcgcatgattgtacc 3165
Db 57 CGCAGCCGCGGAGGTGGGCTCCTGCTCATCTAGGCATCGCACTGATTGTTACC 3

RESULT 15
AAK92794/S
ID AAK92794 standard; cDNA; 546 BP.
XX
AC AAK92794;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human cDNA 3'-end sequence, SEQ ID NO: 1254.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
WPI; 2001-524255/58.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 3; SEQ ID NO 1254; 1380pp + sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesising the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is the nucleotide
CC sequence of the 3'-end of a cDNA provided in the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.

```

```

XX
SQ Sequence 546 BP; 135 A; 127 C; 139 G; 137 T; 8 other;
Query Match 6.9%; Score 480; DB 22; Length 546;
Best Local Similarity 94.0%; Pred. No. 3.1e-121;
Matches 517; Conservative 0; Mismatches 28; Indels 5; Gaps 2;
QY 6125 aactgtcgaagagatgtttaaacctccatagagagggccgagggcgaggggscatctgtgt 6184
Db 545 AACGCCNCAANAATGCTTAAACCCNCCATANA-AGGACCCGCGGCNCGAGGCATTTGTGTG 487
QY 6185 gtcccgctcacacactcagcaggaaggtggctggctgctccctgcatcagatgattt 6244
Db 486 CCNGTCAACATCAGCA----GGAAAGGGTGTGGCTGGCTCCCTGGCATCAGTGGTTT 431
QY 6245 ggtttaagctccagaggggtcttattgcaattgtcttttctctgcccccttgagcagcct 6304
Db 430 GGTTTAAGCTCCAGAGGGTCTTATTGCCATTGTCTTTTCCINTGCCCTTGAGCCAGCCT 371
QY 6305 aaggccctgagctgtgtttttaggcggatgaactgacatgacctcctaccatgaccaggc 6364
Db 370 AAGGCCCTGGAGTCTGTTTCTTTAGCGGATGAACATGACATGCTCTTACCATGACCAGGC 311
QY 6365 tctgggcaagctcctcacagtatccttgagaggtggcagtggaagtgccatttctcag 6424
Db 310 TCTGGGCAAGGCTCCTCACAGTATCCTTGAGAGGTGGCATGGAAGTGCCCATTTCTCAG 251
QY 6425 gtacagaaaccttcagagaggataaatagtcttgccctgtagaagcaggaactgaaacctt 6484
Db 250 GTACAGAAACCTTCAGAGAGGATAAATAGCTTGCCCTGTAGAACGAGACTGAAACCTT 191
QY 6485 gtccgctgactccccagctactctgcccactgtgcccactgtagcccccttactgtcctggca 6544
Db 190 GTCCGCTGACTCCCCAGCTACTCTGCCACTGTAGCCCCCTGCTTACTGTCTCTGGCA 131
QY 6545 caccctcaccatcctgtatataccttaaatcaaaagaggaagcaagagaaagggctttaa 6604
Db 130 CACCCCTCACCATCTCTGTATACCTTAATATFCAAGAGGGGCAAGAGAAAGGGCTTTAA 71
QY 6605 agataagttattttttaaggaaaccttaataattattttaagaagtaaccaaatagtaga 6664
Db 70 AGATAAGTTATTTTAAAGGAACCTTAATATTATTATTTTAAAGAGTAACCAAAATTAGTGA 11
QY 6665 cgtgaaatgc 6674
Db 10 CGTGAATGC 1

```

Search completed: July 26, 2002, 01:51:43
Job time: 20762 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 18:08:11 ; Search time 6874.72 Seconds
(without alignments)
13648.677 Million cell updates/sec

Title: US-09-840-746-2

Perfect score: 6952

Sequence: 1 gttcgtatgaagaattgcgc.....tactgggaagctatagggtg 6952

Scoring table:

IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmd:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_estl:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	698.2	10.0	900	10 BE741056	BE741056 601594040
2	678	9.8	856	10 BG400390	BG400390 602464410
3	650.6	9.4	805	10 BE379009	BE379009 601236278
4	648.8	9.3	745	10 BE876909	BE876909 601488479
5	639.6	9.2	953	10 BG034314	BG034314 602302393
6	631	9.1	745	10 BI223426	BI223426 602943021
7	628.2	9.0	645	10 BI545262	BI545262 603187386
8	625.6	9.0	671	10 BE394218	BE394218 601311526
9	601.4	8.7	630	9 ALS96844	ALS96844 DRZP451P
10	600	8.6	715	10 BE378473	BE378473 601236457
11	546.8	7.9	566	9 AA121502	AA121502 zk88a11.s
12	542.2	7.8	856	10 BF179545	BF179545 601807025
13	541.8	7.8	622	9 AA127811	AA127811 zk88a11.f
14	540.4	7.8	740	10 BI090766	BI090766 602855763
15	528	7.6	528	10 BI714757	BI714757 ic31e03.x
16	526	7.6	527	9 AW192581	AW192581 xl46n09.x
17	520.4	7.5	537	9 AI418494	AI418494 tg48f07.x

18	519.6	7.5	912	10 BE515103	BE515103 601236187
19	515.4	7.4	843	10 W28968	W28968 54d6 Human
20	511.8	7.4	605	9 AW862970	AW862970 MR2-SN000
21	506.8	7.3	523	10 BI548253	BI548253 603191115
22	498.6	7.2	522	9 AW294116	AW294116 UI-H-BI2
23	496	7.1	499	10 BI715315	BI715315 ic31e03.y
24	476	6.8	476	10 BF058997	BF058997 7k36b08.x
25	475	6.8	475	10 BF058131	BF058131 7k38b08.x
26	469.2	6.7	474	10 BE393056	BE393056 601309103
27	467.4	6.7	488	10 BF750474	BF750474 RCI-BN041
28	462	6.6	462	9 AI653145	AI653145 wb43e05.x
29	457.6	6.6	708	10 W22070	W22070 61B5 Human
30	457	6.6	457	9 AW275046	AW275046 xm54q01.x
31	455	6.5	496	10 BF750527	BF750527 RCI-BN041
32	454.8	6.5	458	9 AI829334	AI829334 wk60f12.x
33	454	6.5	454	10 BF724576	BF724576 bx06d11.y
34	452.8	6.5	467	9 AW207638	AW207638 UI-H-BI1-
35	449.8	6.5	453	9 AI754021	AI754021 cr18f10.x
36	449.2	6.5	460	9 AI242664	AI242664 q118b03.x
37	448	6.4	448	9 AI127628	AI127628 qc30f03.x
38	445.2	6.4	462	9 AI143494	AI143494 qa43907.s
39	432.4	6.2	434	9 AI624878	AI624878 ts72e01.x
40	428	6.2	463	10 BF154764	BF154764 PM2-BT072
41	426	6.1	426	9 AA548961	AA548961 nk51e09.s
42	424	6.1	424	9 AA600099	AA600099 ag34q08.s
43	423.4	6.1	425	9 AA912014	AA912014 og22a09.s
44	423	6.1	429	10 BF724575	BF724575 bx06d11.x
45	415.2	6.0	570	10 W52003	W52003 zc92e10.rl

ALIGNMENTS

RESULT 1

BE741056

LOCUS

DEFINITION 601594040F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3948153 5', mRNA EST 15-SEP-2000

ACCESSION BE741056

VERSION BE741056.1 GI:10155048

KEYWORDS EST.

SOURCE human.

ORGANISM

REFERENCE NIH-MGC http://mgc.nci.nih.gov/.

AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE Unpublished (1999)

JOURNAL Contact: Robert Strausberg, Ph.D.

COMMENT Email: cgapbs-x@mail.nih.gov

Tissue Procurement: DCTD/DTP

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Cloned through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLC810 row: P Column: 10

High quality sequence stop: 766.

Location/Qualifiers

1..900

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3948153"

/lab_host="DH10B (phage-resistant)"

/tissue_type="adenocarcinoma cell line"

/notes="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming, Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).".

BASE COUNT	245 a	216 c	239 g	200 t
ORIGIN				

Query Match 10.0%; Score 698.2; DB 10; Length 900;
Best Local Similarity 91.6%; Pred. NO. 3.4e-134;
Matches 796; Conservative 0; Mismatches 63; Indels 10; Gaps 5;

QY 2788 aggatgcagaaatgtgtcaactcctgcaagtcctctgtgtgaggctgtgccagctcttggga 2847

Db 1 AGGATGCAGAAATGTCACACTCTCGAAGTCCTCTGCTGAGGTCGCCAGCTCTGGGA 60

QY 2848 tctcagagcggtatcttttagagcgggcagcttctgtcagcgggaagaggtcccgaatgtgac 2907

Db 61 TCTCAGAGCGGATCTTTA - AGCGGGCAGCTTGTGCAACGGGAAGAGTCCCGAATGTGAC 119

Qy 2908 aaagacactccatctgcactgacctgacggcgctgcctgtgcagtgcaagtcgga 2967

Db 120 AAAGACACCTCCATCTGCACCTGACCTGGACGGCGTTGCCCTGTGCCAGTGCAGTCGGGA 179

Qy 2968 tactttcaggttcacaagaatggaccactcctgccgagcatgtgaagatggatataggctt 3027

Db 180 TACTTTCAGTTCACAAGATGGACCACCTCCGCCGAGCATGTGAAGATGGATATAGGCTT 239

QY 3028 gaaatgaaacctgcatgagttgccatttggccttggtggtctcaactgtgaaaccc 3087

Db 240 GAAATGAAACCTGCATGAGTTGCCCTTGGCCTTGGTGGTCTCAACTGTGGAACCCC 299

Qy 3088 tatcagcttatcactgtggtgatcgccagcgcgagggtgggtcctcctgctcatcctaggc 3147

Db 300 TATCAGCTATCACGTGTGGTATCGCAGCGCGGGAGTGGGCTCCTGCTCATCTAGGC 359

QY 3148 atcgcactgattgttacctgttgagaaaataaaaaatgacataagcaaacatcttc 3207

Db 360 ATCGCACTGATTGTACCTGTTGCAGAAAGAATAAAAATGACATAGCAACATCATCTTC 419

Qy 3208 aaaagtggagatttccaaaagtcccccatatgctgaatacccccaaaaatcctcgctcaca 3267

Db 420 AAAAGTGAGATTTCAAATGTCCCGTATGCTGAATACCCCAAAATCTCGTCAAA 479

Qy 3268 gaatggggccgagaagctattgaaatgcattgagagaatgggaagtaccaaaaacccctctccag 3327

Db 480 GAATGGGGCCGAGAACTATTGAATGCATGAGATGGAAGTACCAAAACCTCTCCAG 539

Qy 3328 atgacggatgtgtactactcgcctacaagtgttaagggaatccagaaacttgacgaaacgga 3387

Db 540 ATGACGGATGTGTACTACTCGCCTACAAGTGTAAAGGATCCAGAACTTGACGAACGGA 599

Qy 3388 ctctaccggcctacactggactgccaggatcacggcaattcttgcattttccccggacag 3447

Db 600 CTCTACCGG-CTACACTGGACTGCCAGGATCAGCGCATCTTGGCATTTTCCCGGACAG 658

Qy 3448 tat-aaccggtctttcatcagtgatgaaagcagaagaagagactactcttttaagtccagga 3506

Db 659 TATAACCGTCTTTTCATCATGATGAACGAGAGAGACTACTTTTAAAGTCCAGG 718

Qy 3507 gagaga---gggactcaattgctctgagcagtcacctcctgggacctctgctcagaggaccgc 3563

Db 719 GAGAGAAAGGGACTCAATTGCTCTGAGCAGTCACTGGGAAC----TCTGTCAAGGACG 774

Qy 3564 accaggaggctgcgccaggatttctcgggagccacgctgagtcgcaagcaggaaagagg 3623

Db 775 CACAGGAGGCTGGGCCCAAGATTGTGGAGGCCAAGCTGAGTGCAGCACGAAAGAGG 834

Qy 3624 acaggcatcgggcggtgaccacagtggg 3652

Db 835 ACACGCTTGGGGGTGACACAGTGGTGA 863

RESULT 2
BG400390

2

1

LOCUS	BG400390	856 bp	linear	EST 12-MAR-2001
DEFINITION	602464410f1 NTH_MGC_75 Homo sapiens cdna IMAGE:4592570 5', mRNA sequence.			

ACCESSION	BG400390
VERSION	BG400390.1 GI:13293838

KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 856)

AUTHORS NIH-MGC <http://mgc.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.

cdna Library Preparation: CLONTECH Laboratories, Inc.
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

/note="Organ: uterus; Vector: POTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGCAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)." BASE COUNT 214 a 209 c 197 g 185 t

ORIGIN

Query Match 9.4%; Score 650.6; DB 10; Length 805;
 Best Local Similarity 97.0%; Pred. No. 2.5e-124;
 Matches 684; Conservative 0; Mismatches 19; Indels 2; Gaps 2;

```

QY 1 gttogatgaaagaattgcgcgtttttcaacaaaagtgtgaaacagcctctggagatgggaac 60
DB 95 GTTCGATCAAAAGAAATTGCCGCTTTTCAAAACAAAGAGTGGAAACAGCCTCGGAGATGGGAAC 154

QY 61 agagaggcgatgggctgtcagaagaatgactgtgcacagcaagaaggccacacttc 120
DB 155 AGAGAGGCGGATGGGGCTGTGAGAAGAAATGACATGTGCACAGCCAAAGAGGCCACCACTTC 214

QY 121 ggcctggagccctctcttctctctgttgagatgggagagctgcaccagccttctag 180
DB 215 GCCTTGGAGCCGCTCTTCTCTGCTTTGGAGATGGGAGAGCTGACCAGCCCTTCTAG 274

QY 181 gaagagaaattcctcaggaccagatctctctggtcgtcatttctacaggacagcagcttc 240
DB 275 GAAGAGAAATTCCTCAGGACCAGATCTCTCTCGTGCATTTCTACAGGACAGCAGCTTC 334

QY 241 ctctctctttagaccttctcactctctgaaagtacagagaagcttaacaactccac 300
DB 335 CTCTCCTCTTTAGACCTTTCTCTATCTTTGAAAGTACAGAAAGCTTAAACAACCTCAC 394

QY 301 tggcctccagagctcctcagtcagtcacaacaaagaacatgcattgtgtaccgtgttcac 360
DB 395 TGGCCTCCAGAGCTCCTCAGTCAGTCAAAACAAAGACAAATGCATGTTGCCACCGTGTTCAC 454

QY 361 tgatggtggccgagaaacgctgcgactctttgacggtcagctctgggaacctgtgagcaagac 420
DB 455 TGATGGTGGCCGAGAACGCTGCGATCTTTGACGGTCAGTCTGGGACCTGTGTGAGCAAGAC 514

QY 421 aagagcttccccaaagagactccagaattggccagacttcattcctcagctctctttcacc 480
DB 515 AAGAGGCTTCCCAGAGGACTCCAGAATTTGCCAGACTTCATCCTCAGTCTCTTTTTCACC 574

QY 481 ctctcagtggaatcgagaagaaacagtagagtaactgggaattccagggggatgaggaatt 540
DB 575 CTCGCACTGGAATCGAGAAGAAACAGTAGAGTAAC TGGGAATCCA -GGGATAGGAAATT 633

QY 541 cattgaaccatccacagaaaaatgaattggacttacgtcttttgcgtggcaaaatgatctcc 600
DB 634 CATTGAACCATCCACAGAAAAATGAATTGGACTTACGTCTTTTGGCGTTGGCAAAATGATTCC 693

QY 601 ccaacc-tttgagaacatcagcttgcagcagctctgagtgcaaaatgaagatcccat 659
DB 694 CCAACCTTTTGGGAACATCAGCTTGGCAGCAGCTCTGTAGGTGCAAAATGGAAGTCCCAT 753

QY 660 gtctcagactgagactgtgtctaggtcagtcgcaccccatgagag 704
DB 734 GTCTCAGACTGAGACTGTGTCTAGGTGAGTCGCACCCCATGAAAG 798
  
```

RESOLUTION	4
BE876909	
LOCUS	745 bp mRNA linear EST 20-OCT-2000
DEFINITION	601488479f1 NIH_MGC_69 Homo sapiens CDNA clone IMAGE:3890899 5', mRNA sequence.
ACCESSION	BE876909
VERSION	BE876909.1
KEYWORDS	GI:10325672
SOURCE	EST, human.

ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 745) NIH-MGC http://mgc.nci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-remail.nih.gov Tissue Procurement: DCTD/BTP/Gazdar cDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM9674 row: n column: 20 High quality sequence stop: 644. Location/Qualifiers 1. 745 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:3890899" /clone_lib="NIH_MGC_69" /tissue_type="large cell carcinoma, undifferentiated" /lab_hosts="DHI0B (phage-resistant)" /note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; cloned unidirectionally. Primer: Oligo dt. Average insert size 1.1 kb. Library constructed by Life Technologies." 174 a 198 c 217 g 156 t
FEATURES	
source	
BASE COUNT	
ORIGIN	

Query Match	9.3%	Score 648.8	DB 10	Length 745
Best Local Similarity	97.4%	Pred. No. 5.9e-124		
Matches 670	Conservative 0	Mismatches 17	Indels 1	Gaps 1

QY	5838	ggggacatgagtaagaacatgtctgtgacccagtgaggaaaggagagggccagctgcactcc	5897
DB	1	GGGACATGAGTAAGACATGTCGTGTGACCCAGTGAGGAAGGAGAGGCCAGCTGCACATCC	60
QY	5898	tgcacggggttccttagctgcagaaggggtccgcgctaggccgaggggaaacacctgatagc	5957
DB	61	TGCACGGGTTCTTAGCTGCAGAAGGGTCCCGCCTAGGCCGAGGGGAACACCTGTATAGC	120
QY	5958	agaagagccctggtatgcacacctggcaocgcgagggctctccgccagacacagtgctcca	6017
DB	121	AGAAAGGCCCTGGATGACACACTGGCAGCGCGAGGCTCTCCGCCAGACACAGTGTCTCCA	180
QY	6018	tgtcagccctcgacctggggtgtgtattcaogtgcacagatgcccaaatccttcgaccca	6077
DB	181	TGTCAGCCCCCTGCACACTGGGGTGTGTGATTACGTGCACAGATGCCAATCTCTGCACCA	240
QY	6078	atatccacacagatgggggaaggtgagaggaagggcgaagtgatgtgaactgctccaagag	6137
DB	241	ATATCCACAGATGGGGGAAGGTGAGAGGAAGGGCAAGTGATGTGTAACTGCTCAAGAG	300
QY	6138	atgcttaaacctcctaatagagagagccggggcgcagggggcactctgtgtgtccggtcacaca	6197
DB	301	ATGCTTAAACCTCCATAGAGAGGACCGGGCGCAGGGGCATCTGTGTCTCCCGTCACACA	360
QY	6198	ctgcagcaggaaggggtgctgtgctgctcctagccatcagtggttggttaagctcca	6257
DB	361	CTGCAGCAGGGAAGGGTGGCTGGCTGGCTCCTGGCATCAGTGGTGTGGTTTAAGCTCCA	420
QY	6258	gaggggtcttattgccattgtcttttctctgccctctgagccagcctaaagccctggagt	6317
DB	421	GAGGGTCTATTGCCATTGCTTTTCTCTGCCCCCTTGAGGCCAGCCPAAGSCCCTGGAGT	480
QY	6318	ctgtttcttttaggcggatgaactgacatgtctctaccatgaccaggctctgtgggcaaggct	6377
DB	481	CTGTGTTCTTTTAGGCGGATGAATGACATGCTCTTACATATCAGAGGCTCTCGGCAAGGCT	540

Qy	6378	cctcacagtagtacctgagaggtgagcctggaagtgcacatttccaggtacagaaacctt	6437
Db	541	CCTCACAGTATCCTTGAGAGGTGGCGATGGAAGTGCCATTCTCAGGTACAGAAAACCTT	600
Qy	6438	cagagagagataaatagtctgcccctgtgtagaagcagagactgaaacccctgtccgcctgacct	6497
Db	601	CAGAGAGATAAATAGCTTGCTGTAGAAGCGCAGGACTGAAACCTTGTCGGCGCTGTATCC	660
Qy	6498	ccccagctactctgccacactgtagcccc	6525
Db	661	CCCCAGCTA-TCTGACCACCTGTAGCCCC	687
RESULT	5		
LOCUS	BG034314	602302393F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4403831 5',	EST 24-JAN-2001
DEFINITION		mRNA sequence.	
ACCESSION	BG034314		
VERSION	BG034314.1	GI:12427500	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 953)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs-r@mail.nih.gov Tissue Procurement: DCTD/DTM cDNA Library Preparation: Life Technologies, Inc. cDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM10114 row: j column: 24 High quality sequence stop: 614. Location/Qualifiers 1. .953 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4403831" /clone_lib="NIH_MGC_87" /tissue_type="mammary adenocarcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: breast; Vector: pCMV-SPORT6; Site_1: NOTI; Site_2: Sali; Cloned unidirectionally; oligo-dT primed. Average insert size 1.393 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH_MGC Library." 242 a 237 c 286 q 188 t		
BASE COUNT			
FEATURES	source		

	Query Match	9.28;	Score 639.6;	DB 10;	Length 953;
	Best Local Similarity	95.6%;	Pred. No. 4.8e-122;		
	Matches 690;	Conservative	0; Mismatches 29;	Indels 3;	Gaps 3;
Qy	5465	cccgagctgtggaagaccatagtggaggcgtattcccaggcatcacacttccaatgcc	5524		
Db	1	CCCGAGCTGTGGGAAGGCCATGTAGGAGGCATTTC CAGGCATACACTTCCACATGCC	60		
Qy	5525	ttcagctgcagctcacgctgacaaatcatctccctctatcggagcagaagaacttcagctc	5584		
Db	61	TTCAGCTGATGTCACAGCTGACAATCATCTCCTCTATCGGAGCCAGAGACTTCAGCTC	120		
Qy	5585	cacaaaatgaagtgtctgtcctgataaacatctcttggaagaatccccacaatcagaaaaa	5644		
Db	121	CACAAATGAAGTGTTCTGTCTGAAAACATTCTTGGGAAGAATCCC AAATCCGAAAAA	180		


```

DEFINITION 603187386F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5259020 5',
mRNA sequence.
ACCESSION BI545262
VERSION BI545262.1 GI:15432574
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 645)
NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LL4M1452 row: k column: 21
High quality sequence stop: 645.
FEATURES
source
1..645
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5259020"
/clone_lib="NIH_MGC_95"
/tissue_type="hippocampus"
/lab_host="DH10B"
/note="Organ: brain; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTNN-3',
size-selected for average insert size 2.5 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 152 a 149 c 183 g 161 t
ORIGIN

Query Match 9.0%; Score 628.2; DB 10; Length 645;
Best Local Similarity 99.4%; Pred. No. 1.1e-119;
Matches 641; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 3475 agcagaagaagagactacttttaagtccaggagagagaggactcattgctctgagccag 3534
DB 1 AGCGGAGGAAGAGACTACTTTTAAGTCCAGGAGAGAGAGGGACTCATTCGCTGAGCCAG 60

QY 3535 tcacctggagacctctctcagagagaccgaccaggaggctgcgccagaga-tttgtcg99 3593
DB 61 TCACCTGGGACCTCTCTCAGAGAGACCGCACAGGAGGCTGCCGCCAGGATTTGTGGG 120

QY 3594 agccacgctgagtggcaagcaggaagaggagagcagcatcggggctgaccacagtgag 3653
DB 121 AGCCACGCTGAGTGGCAAGCAGGAAGAGGACAGGACATCGGGGGCGCTGACCAAGTGGAG 180

QY 3654 gagacaggttgatgtggaaccacagcgtctcattcagcacctttgttactgtgaac 3713
DB 181 GAGACAGGTGGATGTGGAAACCACAGCGTCTCATTCAGCACCTTTGTGTACTGTGAAC 240

QY 3714 gtgaatgtgggccaagtatacaagagagctctctctgagtgcactgcaccatggcactggcacc 3773
DB 241 GTGAATGTGGGCGAGTATCAAGAGAGTCTCTCTGAGTACTGCAACCATGCGACCTGGCACC 300

QY 3774 agggcgactattagccaggcgagaccactagactcttcagtcaggagacctggtttccctt 3833
DB 301 AGGGCGACTATTAGCCAGGTGAGACCACCTAGACTTTTCAGTGCAGGGACCTGGTTTTCCCTT 360

```

```

QY 3834 cgttgcacttaagtaaatgggtggaggtttctcttttgatctgttttgagactgttc 3893
DB 361 CGTTGCACCTTTAGTAATTTGGGTGGGAGGTTTCCCTTTTGATCTGTTTGGAGACTGTC 420

QY 3894 cagaagaagagcttccttcccagagacacattccatagcagcaaatttgggtgattcattg 3953
DB 421 CAGAAAGAAGGCTTCCCTTTCCCGAGAGACATTCATAGCAGCAATTTGGTGATTCATTG 480

QY 3954 cagcaaaataactggcttgtaattatttctctgcccagcgctgcgtgctaaacaaga 4013
DB 481 CAGCAAAATACTGGCTTGTTAATTTATTTTCTTGCCCCAGCGCTCGTCTAAACAACAGA 540

QY 4014 tgaggatgagcgtaccacactgaagtctgaagatgctgcattgaacgacagtgtttcat 4073
DB 541 TGAGGATGACGGGTACCACTGAAGTCTGAAGATGTGCCATTTGAACGACACAGTGTTCAT 600

QY 4074 atgtttcaggtgtcttatgctacagtttcccaagccagcccca 4118
DB 601 ATGTTTCTAGTGTCTCTTATGCTACAGTTTCCAAAGCCAGCCCCCA 645

RESULT 8
BE394218 BE394218 671 bp mRNA linear EST 21-JUL-2000
LOCUS 601311526F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3632959 5',
mRNA sequence.
ACCESSION BE394218
VERSION BE394218.1 GI:9339583
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 671)
NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE Unpublished (1999)
JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM325 row: c column: 08
High quality sequence stop: 661.
FEATURES
source
1..671
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3632959"
/clone_lib="NIH_MGC_44"
/tissue_type="endometrium, adenocarcinoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming, Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 160 a 179 c 184 g 148 t
ORIGIN

Query Match 9.0%; Score 625.6; DB 10; Length 671;
Best Local Similarity 99.0%; Pred. No. 3.8e-119;
Matches 661; Conservative 0; Mismatches 4; Indels 3; Gaps 3;

QY 5325 ctctctctgtagaaccaatggccctttgtggtctcacggcctgcacctaactggagag 5384

```

```
Db 1 CTCTCTCTCTGTAGAACCAATGGCCCTTTGTGGCTCAGGCCTCGCACCTCAACTGGAGAG 60
QY 5385 ttctgaactctcagctacactgagccacagactaggcttcttgctccttcctccgagc 5444
Db 61 TGCTGAGCTCTCTGAGCTCACTGAGCCACAGACTAGGCTTCTTGCTCTCTCCGAGC 120
QY 5445 atgctctctaccccccagaccgacgctgtggaaagacatgtaggagctatcccc 5504
Db 121 ATGCCTGCTCACCCTCCAGACCCGACACTGTGGAAAGAGCCATGTAGGAGGCTATTCCC 180
QY 5505 aggcataacttccactgccttcagc-tgacgtcacagctgacaaatcatctcctctatc 5563
Db 181 AGGCATACACTTCCTACCTGCTTACGCTGACGTACAGCTGACAAATCATCTCCTCTATT 240
QY 5564 ggagccagaagacttcagctccacaaatgaagtgtctctcctgaaacattctggga 5623
Db 241 GGAGCCAGAGACTTCAGCTCCACAAATGAAGTGTCTGCTGTAAGAAATCTTGGGA 300
QY 5624 agaatcccaacatcgagaaacgggtgcctgtgagttcccaaatgcttctgttctatgg 5683
Db 301 AGAATCCACATCGAGAAACGGGTCTGTGAGTTCCAAACATGCTTCTGTTCATGG 360
QY 5684 gttcttccctatggagtgattaaagagtggtttttattttgttcttaactgagaaaaa 5743
Db 361 GTTTCTTCGATGAGTGGATTAAAGAGTGTTTATTATTGTGTTCTTAAGTGAAGAAA 420
QY 5744 agagggcaccacaaaggttgaggtcacacagttccacagtttccagaggcggttgggg 5803
Db 421 AGGAGGCCACCCACAAGGTTGAGGTTCACACAGTCTCCACAGTTTCCAGAGGGCG-TTGGGG 479
QY 5804 gtggggagagcactcagacatgaagctcctaaggagacatgagtaaacatgtctgtg 5863
Db 480 GTGGGGAAGGCACCTCCAGAGCATGAGGCTCTAAGGGGACATGAGTAAGCATGCTGTG 539
QY 5864 accagtgagaaaggagagggcagctgcactcctcctgcacggggttcttagctgcagaag 5923
Db 540 ACCAGTGAGAAAGGAGAGGCCAGCTGCACCTCTGACGGGTTCTTAGCTGCAGAGG 599
QY 5924 gtcccgctaggcaggggaaacacctgatagcagaagaggcctggtgatcacacctggc 5983
Db 600 GTCCCGCTTAGCCGAGGGGAAACACCTGTATAGCAG-AGAGGGCTGGATGCACACCTGGG 658
QY 5984 acgcccag 5991
Db 659 ACGCCGAG 666
```

```
RESULT 9
AL596844 LOCUS
DEFINITION DKF2p451p091_r1 451 (synonym: hlcc1) spinal cord Homo sapiens cDNA
clone DKF2p451P091 5', mRNA sequence.
ACCESSION AL596844
VERSION AL596844
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 630)
AUTHORS Poustka,A., Wellenreuther,R., Mewes,H.W., Weil,B. and Wiemann,S.).
TITLE EST (Poustka,A., Wellenreuther,R., Mewes,H.W., Weil,B. and Wiemann
,S.)
JOURNAL Unpublished (1999)
COMMENT Contact: Poustka A.J.
Department Leirach
Max-Planck-Institute for Molecular Genetics
Innestrasse 73, 14195 Berlin, Germany
Tel: +49-30-84131623
Fax: +49-30-84131128
Email: poustka@mpimg-berlin-dahlem.mpg.de
This is the 5' sequence of the clone insert
```

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project.
No s1 sequence available.
This clone (DKF2p451P091) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
1. 630
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKF2p451P091"
/clone_lib="451 (synonym: hlcc1) spinal cord"
/tissue_type="human spinal cord"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT 150 a 201 c 93 g 186 t
ORIGIN
Query Match 8.7%; Score 601.4; DB 9; Length 630;
Best Local Similarity 99.8%; Pred. No. 4e-114;
Matches 602; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1057 ctgtccattacagataaacagttcctcagacattcctcttcttctcagtcagacagagaga 1116
Db 28 CTGTCCATTACAGATAACAGTTCATCTCAGACATTTCTGGAGAGCTCAACTTCTTATATT 87
QY 1117 aaaaatcacaactcttcacattcagagattcctctcttcttctcagtcagacagagaga 1176
Db 88 AAAATCTCAAACTCTTCACACTCAGAGTATTCCTCTCTTCATGTCTCAGACTGAGAGA 147
QY 1177 agtaaatctcactctatgacggggaatagctcagccttctcactagtcagtcagtcgtg 1236
Db 148 AGTAACATCTCATCTATGACGGGGAATATGCTCAGGCTTCTACTGAGTCGCCAGTCTG 207
QY 1237 catacatccaaaccttcctgctcctacacacccacccaatataatgccgaacaccttcgtgtt 1296
Db 208 CATACATCCAACTTCCTGCTCTACACACCCACCACCTTAATATATGCGGAACACTTCGGTGT 267
QY 1297 ctggaaactgatgctgagttgttagtgactcctcctcctcctcctcctcctcctcctct 1356
Db 268 CTGGACACTGATGCTGAGTTGTAGTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 327
QY 1357 tcttctcttcagggctcctttgctcctctgctcctctgttcacaaatccacattttatt 1416
Db 328 TCTTCTTCTTCAGGGCTCTCTTGGCTCTGCTCTGTCACAAATCCACCACTTTATTT 387
QY 1417 tcataaatattaccatcaaacaggcctctgtgcatctactactaaagtctactctgtatgca 1476
Db 388 TCATCAATTTTACCATCAACACGGGCTCTGTGCATCTACTATAAGTCTACCTCTGATGCA 447
QY 1477 tcaacacatggcttctcctcacaatcacttaccagatactccttaacagacatctacatct 1536
Db 448 TCCACACATGGTCTTCTCTCACCATCACCCTTTACCAGTATCTTTAACGACATCTACATCT 507
QY 1537 gccccatttctgtctcacaaaacaccttgccacagtcattcttaccctctcctgctccc 1596
Db 508 GCCCCTCTTCTGTCTCACAAACAACTTGGCAGAGTCACTTCTACCCCTGTCTCTGCC 567
QY 1597 agggcaaggagagactcctgtgacttcttcagacatcaacaatgacatcattcatgaca 1656
Db 568 AGGCAAGGAGAGACTCTCTGTGACTTTCATTTCAGACATCAACAATGACATCATTCATGACA 627
QY 1657 atg 1659
Db 628 ATG 630
RESULT 10

Qy	6187	ccgcctcacacactgcagcagggaagggtggctggctggctccctcgcatcagtggtttg	6246
Db	361	CCCGTCACACTGCAGCAGGGAAGGGTGGCTGGCTCCCTGGCATCAGTGGTTGG	420
Qy	6247	ttaagctccagaggggtcttatgtgcaattgtcttttctctgccccttgagccagcctaa	6306
Db	421	TTTAAGCTCCAGAGGGGTCTATTATGCCAATTGCTTTTCTCTGCCCCCTTGAGCCAGCCTAA	480
Qy	6307	ggcctggagctgtcttcttaggcggatgaactgacatgctctctaccatgaccaggctc	6366
Db	481	GGCCCTGGAGTCTGTATCTTTAGCGCGATGACTGACATGCTCTTACCATGACCAGGCTC	540
Qy	6367	tggg-caaggctctcacagtatctctgagagggtggcagtg-ccatttctcag	6424
Db	541	TGGCCCAAGGCTCTCACAGTATCCTTGAGAGTGGGCATGGAAGTGGCCCATTTCTCAG	600
Qy	6425	gtacagaaacctcagagagataaatagcttgccctgtagaagcaggactgaacctt	6484
Db	601	GTACAGAAACCTTCAGAGAGGATTAATAGCTTGCCTGT---AGACGAGGACTGAACC--T	655
Qy	6485	gtccgctgactccccagactactctgccactgtagccc	6524
Db	656	TGTCGGCTGACTCCCC--AGTAATCTGCCACTGTAGCCC	693
RESULT	11		
AA121502/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			

BASE COUNT	143 a	138 c	142 g	141 t	2 others
ORIGIN					
Query Match	7.9%; Score 546.8; DB 9; Length 566;				
Best Local Similarity	99.1%; Pred. No. 8.3e-103;				
Matches	559; Conservative 0; Mismatches 4; Indels 1; Gaps 1;				
Qy	6111	ggcaagtgatgttaactgtctcaagagatgcttaaaacctccatagagagagcgcg	6170		
Db	563	GGNCAATGATGTGTAACTGCTCAAGAGATGCTTAAACCTCCATAGAGAGGCGGCGC	504		
Qy	6171	agggcaatctgtgtccctcaacactcagcaggaagggtgctggtggtccct	6230		
Db	503	AGGGCATCTGTGTCTCCNGTCAACACTGCAGCAGGAAGGTGGCTGGCTCCCT	444		
Qy	6231	ggcatcagtggttgggttaagctccagaggtcttattgccattgtctttccctgcc	6290		
Db	443	GGCATCAGTGTGTGTTTAAGCTCCAGAGGGTCTATTGGCATTTGCTTTCTCTGCC	384		
Qy	6291	ctttgagccagcctaaggccctggagctgtttcttttagcggatgaactgacatgctcc	6350		
Db	383	CCTTGAGCCA-CCTAAGGCCCTGGAGTCTGTTCTTTAGCGGATGAACCTGACATGCTCC	325		
Qy	6351	taccatgaccaggtctctggcaaggtccctcacagtatccttgagaggtggcatggaag	6410		
Db	324	TACCATGACCAAGGCTCTGGGCAAGGCTCTCACAGTATCTCTTGAGAGGTGGCATGAAG	265		
Qy	6411	tgccattcttcaggttacagaaaccttcagagaggaataatagcttgccctgtagaagca	6470		
Db	264	TGCCCATTTCTCAGGTACAGAAACCTTCAGAGAGGATAAATAGCTTGCCTGTAGAAGCA	205		
Qy	6471	ggactgaaacctgttcgctgactccccagctactctgcccactgtagccccctgcc	6530		
Db	204	GGACTGAACCCCTGTGCGCCCTGACTCCCCAGCTACTTGCACACTGTAGCCCCCTGCC	145		
Qy	6531	ttactgtcctggcacacctccaccatcctgtatataccttaataatacaagaggcaagag	6590		
Db	144	TTACTGTCTGGCACACCCCTCACCATCTGTATACCTTAAATATCAAAAGGGCAAGAG	85		
Qy	6591	agaaagggctttaaagataagttattttttaaggaaaccttaataatttttaagaagt	6650		
Db	84	AGAAAGGCTTTAAAGATAAGTTATTTTAAAGGAACCTTAATATTATTTTAAAGAAGT	25		
Qy	6651	aaccaaattagtcagctgaaatgc	6674		
Db	24	AACCAAAATTAGTACGTGAATGC	1		
RESULT	12				
LOCUS	BF179545				
DEFINITION	601807025F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4037671 5', mRNA linear EST 31-OCT-2000				
ACCESSION	BF179545				
VERSION	BF179545.1				
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
AUTHORS	1 (bases 1 to 856)				
TITLE	NIH-MGC http://mgc.nci.nih.gov/				
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)				
COMMENT	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgapbs-r@mail.nih.gov				
	Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys				
	cDNA Library Preparation: Life Technologies, Inc.				
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)				
	DNA Sequencing by: Incyte Genomics, Inc.				
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:				

http://image.llnl.gov					
Plate: LLM9315	row: b column: 08				
High quality sequence stop: 713.					
Location/Qualifiers					
source	1. .856				
	/organism="Mus musculus"				
	/strain="C57/B6"				
	/db_xref="taxon:10090"				
	/clone="IMAGE:4037671"				
	/clone_lib="NCI_CGAP_Mam5"				
	/tissue_type="tumor, gross tissue"				
	/dev_stage="7 months"				
	/lab_host="DH108"				
	/note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: Sali; Site:2: Noti; Cloned unidirectionally. Primer: Oligo dr. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH"				
BASE COUNT	232 a	223 c	221 g	180 t	
ORIGIN					
Query Match	7.8%; Score 542.2; DB 10; Length 856;				
Best Local Similarity	83.3%; Pred. No. 7.5e-102;				
Matches	652; Conservative 0; Mismatches 128; Indels 3; Gaps 3;				
QY	2710	gagtcacacgggtgtgtatctcaactgcaaacacccctttccctggcctccaaatgtgacg	2769		
Db	1	GAGCCAGTAGTACAGTTTTCATCTCCTCACTGAAACACCACTTTGCTGGCTGCAATGTGACA	60		
QY	2770	ctatttgacctgctgtagatgcagaatgtgcaactcctgcagctcctgctgag	2829		
Db	61	CTGTTTGACCTGGCAGACAGGATCCAGAAATATGTCAACTCTCTCAGGTCTCTGCTGAA	120		
QY	2830	gtctgacagctcttggatctctcagagcggtatcttttagagcggcagcttgtgcaagcgg	2889		
Db	121	GTCTGCCAGCTCTTGGGGTCTCAGAGCGGGTCTTTTCGAGGGGCAGCTTGTGTAGCGG	180		
QY	2890	agagtcctccgaatgtgcagaagacacctccatctgcactgacctggacggcggttgcctg	2949		
Db	181	AAGATCCAGAATGTGACAAAGAGACCTCCATCTGCAACCGACCTGGATGTTGTCGCTC	240		
QY	2950	tgccagtgcaagtcgggatactttcagttcaacaagatgaccactcctccagagcatgt	3009		
Db	241	TGTCAGTGCAGAGTCTGGCTACTTCCAGTTTCAACAGATGGATCACTCTTCCCGAGCATGT	300		
QY	3010	gaagatggatataggcttgaaaaatgaaacctgcagtgagttgcccatttggccttgggtggt	3069		
Db	301	GAAGATGGATATAGGCTTGAAATGAAACCTGTATGAGTTGCCCATTTGGCTCGGTGGA	360		
QY	3070	ctcaactgtgaaacccctatcagcttatcactgtggtgacgcagcgcgggaaggtggg	3129		
Db	361	CTCAACTGTGGAACCCCTATCAGTCACTCACTGTGTGATGCGCAGCGGAGGTGGC	420		
QY	3130	ctctgctcatcttagcctgcagctgattgttacctgttgacagaagaataaaaaatgac	3189		
Db	421	CTCTGCTCATCTCGCGGTGGCAGCTCATTTGTTACCTTGTTCAGAAAGAGTAAAAATGAC	480		
QY	3190	ataagcaaacctatcttcaaaagtggagatttccaaatgttccccaatgtgtgaatacccc	3249		
Db	481	ATAAGTAAACTCATCTTCAAAAGTGGGAGCTTCCAAATGTCCCATACAC-GATGTCCCC	539		
QY	3250	aaaaatcctcgtcacagaatggggccggaagctattgaaatgcattgagaatggaagt	3309		
Db	540	AAGAACTCTCGTCGCAAGAGTGGGGCCGAGAGCCATTGAAATGCGAGAGACGGAAGC	599		
QY	3310	accaaaacctctccagatgcaggtgtgtactactcgcctacaaagtgaagaatccca	3369		
Db	600	ACCAAAAACTCTCGAGATGACAGACCGTGTACTACTCCGCACAAACGTAAAGAACCTC	659		
QY	3370	gaacttgaa-cgaacgggactctaccggcctacactggactgaggaatccagggcatc	3428		
Db	660	GAACTTGAACCGCAAGGACTGTAAACCAAGCTACACTGGACTGCC-GGATCGAGAACTCT	718		

/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5 kb. Library prepared by Life
Technologies."

```

BASE COUNT      177 a   176 c   185 g   202 t
ORIGIN

Query Match      7.8%; Score 540.4; DB 10; Length 740;
Best Local Similarity 91.2%; Pred. NO. 1.8e-101;
Matches 675; Conservative 0; Mismatches 51; Indels 14; Gaps 9;

QY 4742 aaacgtgttatactctccagtcagtggtggagaaactgaacactccggttggtcgagaa 4801
      |||
Db 1 AAGCTGTTTATCTTCCAGTCAGTCGTGGGAGAACTGAAGACTTCGGTGGTCGAGGAA 60
      |||

QY 4802 ctgagggtgacctctcgggaagaagtcccaactcattattattatgcctgtgatgtg 4861
      |||
Db 61 CTGAGGCTTGACCTTCGGGAAGGAAGTTCACACTCATCTTATTATTATGCTGTGATGTG 120
      |||

QY 4862 ggtcctgcaggagacatccagtactcgtgtgcttcttaattgccacctgggaactgtgt 4921
      |||
Db 121 GGTCCTGCCAGGAGACATCCAGTACTCGGTGCTTTTAAATTGGCCACCTGGGGAACTGTGT 180
      |||

QY 4922 ttattggcctcttctggggcatcctcgtgttttgatgaagtgaagggaatacagaggttaa 4981
      |||
Db 181 TTATTGGCCCTCTCTTGGGGCATCTCTGTTTGGATGAAGTGAGGGGAATACAGAGGTAAA 240
      |||

QY 4982 agaattgtctccacctgaagcggggagtcctccgcttcacatttctggaatggtgcagcc 5041
      |||
Db 241 AGAATTGCTCCACCCCTGAAGCGGGAGTCCCGCTTCACATTTCTTGGAAATGGTGCAGCC 300
      |||

QY 5042 actggggacagtctgc-cccgggcatggt-tgtttctcaaggctcctcaataataatc 5099
      |||
Db 301 ACTGGGGACAGTCTGCACCCGGGCATGGTATGTTCTTCAAGTCTCTTACGTTATAATC 360
      |||

QY 5100 cctattctcataaactccttg-gccttgatggttttaagcaagaactcctggtgcccatg 5158
      |||
Db 361 CTTATTCTTACATAATCTTGGACCCCTGATGGTTTAAAGCAAGAACTCCTGTGTCCCATG 420
      |||

QY 5159 gtctccaccactcaccatcaccctgtgtgtagcaagatcctagtcaggggaggtgcatt- 5217
      |||
Db 421 GTCTCCAGCACTCACCATCACCTCGTGTGAGCAAGAGTCTCTAGTCAGGGGAGGTGCTATTA 480
      |||

QY 5218 ttagttagtaaatgcaattatcc-atgagataaaataaaggagaactgttt-ttatca 5274
      |||
Db 481 TAAGTAGTTACATGGCACTTATCCAAATGAGATAAATAACAAGGAGAACTGTTACTATCCA 540
      |||

QY 5275 gtgagggttaacctaaattcaaaagtgcgcctttttgaaattcttgggc---ctctctc 5331
      |||
Db 541 GTGGAGGCTAACCTTAAATTTCAAGTGTGCGCTTTTAAAGAACTTGGGGCCATCTCTCTC 600
      |||

QY 5332 tctgtagaaccaatggcccttttgg---ctcacgacctgcacctaac-tggagagttc 5387
      |||
Db 601 TGTAGAAACCAATAGGGCCCTTTGTGGGATCACGGGCGCTGCACACACTAACTTGGAGAGTTC 660
      |||

QY 5388 tgagctcctcagctcaactgagcccaacagactaggcttcttggctccttcgcgagcatg 5447
      |||
Db 661 TGAGCTCCTGTAGCTACCTTGACACACACAGACTATGCTCTTGGCTCCATCCGCAAGCATG 720
      |||

QY 5448 cctgctcaccgccagaaccc 5467
      |||
Db 721 CTTGCTAACCCCAAGACCGC 740
      |||

```

RESULT 15

BI714757/c 528 bp mRNA linear EST 19-SEP-2001
 LOCUS BI714757
 DEFINITION 1c3ie03.x1 HR85 islet Homo sapiens cDNA 3', mRNA sequence.
 ACCESSION BI714757
 VERSION BI714757.1 GI:15690452
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 528)

REFERENCE
AUTHORS

Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
 Lemishka, I., Searce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A.,
 Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas
 M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T.,
 Jackson, Y., and Bowers, Y.
 Endocrine Pancreas Consortium
 Unpublished (2000)

TITLE
JOURNAL
COMMENT

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohp.harvard.edu

Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
 Washington University Genome Sequencing Center for information on
 obtaining a clone please contact: Dr. Hiroshi Inoue
 (hinoue@im.wustl.edu)
 Seq primer: -400P from Gibco
 High quality sequence stop: 446.

FEATURES
source

1..528
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="HR85 islet"
 /tissue_type="Purified pancreatic islet"
 /lab_host="DH10B"
 /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
 NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
 Size-selected on agarose gel. Average insert size ~1kb. 5'
 XhoI site was destroyed after directional cloning.
 Amplified once. Contact information: Hiroshi Inoue, MD,
 Metabolism Div. (Alan Permutt Lab), Washington University
 School of Medicine, Box 8127, 660 South Euclid Ave., St.
 Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
 314-362-1916, Fax: 314-747-2692."

BASE COUNT
ORIGIN

134 a 130 c 137 g 127 t
 Query Match 7.6%; Score 528; DB 10; Length 528;
 Best Local Similarity 100.0%; Pred. No. 6.6e-99;
 Matches 528; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 6148 ctccatagagaggagcggcgaggggcatctgtgtcccgtcacacactgcagcagg 6207
      |||
Db 528 CTCATAGAGAGGAGCGGGCGAGGGGCATCTGTGTCCCGTCACACACTTGCAGCAGG 469
      |||

QY 6208 gaagggtgctggtgcctgcctgcctgcatcagtggttttaagctccagaggtctta 6267
      |||
Db 468 GAAGGGTGGCTGGCTGCTCCTGGCATCAGTGGTTTGGTTTAAGCTCCAGAGGGTCTTA 409
      |||

QY 6268 ttgcatgtcttcttctcctgccttgagccagccctaaaggccctggagctgttctttt 6327
      |||
Db 408 TTGCCATTGTCTTTCTCTGCCCCCTTGAGCCAGCCCTAAGGCCCTGGAGTCTGTTCTTT 349
      |||

QY 6328 agcgagatgaactgacatgctcctaccatgacaggtcctgggcaaggtcctcacagta 6387
      |||
Db 348 AGCGGATGAACATGACATGCTCTTACCATGACCAAGGCTCTGGGCAAGGCTCTCCACAGTA 289
      |||

QY 6388 tccttgagagtggtgcatgaagtgccttctcaggtacagaaacctcagagagat 6447
      |||
Db 288 TCCTTGAGAGTGGGATGGGAAGTGCCCATTTCTCAGGTACAGAAACCTTCACAGAGGAT 229
      |||

QY 6448 aatagcttgcctgtagaagcaggtgactgaaaccttgcgctgactccccagctac 6507
      |||
Db 228 AATAGCTTGGCCCTGTAGAGCAGGACTGAACCCCTTGTCCGCCCTGACTCCCCCAGCTAC 169
      |||

```

```

Qy 6508 tctgccacttagcccccctgcttactgtcctggcacacccctcaccatccctgtatacc 6567
      |||||
Db 168 TCTGCCCACTGTAGCCCCCTGCCTTACTGTCTCTGGCACACCCCTCACCATCCTGTATACC 109
      |||||
Qy 6568 ttaaatatcaagagggggcaagagagaaaggcctttaagataagttattttttaaggaa 6627
      |||||
Db 108 TTAAATATCAAGAGGGGCAAGAGAGAAAGGCTTTAAAGATAAGTTATTTTTTTAAGGAA 49
      |||||
Qy 6628 ccttaatatattttttaagaagtaaaccaaatagtgacgtgaaatgca 6675
      |||||
Db 48 CCTTAATATTATTTTAAAGAAGTAACCAAAATTAGTGACGTGAAATGCA 1
      |||||

```

Search completed: July 25, 2002, 22:06:58
 Job time: 14327 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 25, 2002, 20:03:31 ; Search time 192.67 Seconds
(without alignments)
8863.051 Million cell updates/sec

Title: US-09-840-746-2
Perfect score: 6952
Sequence: 1 gttcgaatgaagattgcgc.....tactgggaagctatagggtg 6952

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCRUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5098.4	73.3	6407	2	US-08-616-844-7
2	5098.4	73.3	6407	2	US-08-599-654-7
3	5098.4	73.3	6407	3	US-08-944-868A-7
4	5098.4	73.3	6407	3	US-08-944-423A-7
5	5098.4	73.3	6407	3	US-08-944-496-7
6	470.4	6.8	584	4	US-09-328-111-83
7	126.4	1.8	128	2	US-08-480-994-7
8	126.4	1.8	128	2	US-08-485-573-7
9	126.4	1.8	128	3	US-08-925-743-7
10	126.4	1.8	128	4	US-08-925-767-7
11	89.8	1.3	7218	1	US-08-232-463-14
12	56.2	0.8	3489	2	US-08-728-323A-1
13	56.2	0.8	32207	2	US-08-770-379-20
14	56.2	0.8	32207	4	US-08-757-669A-20
15	56.2	0.8	32207	4	US-09-230-371A-20
16	52.8	0.8	2518	4	US-09-433-699-3
17	49.4	0.7	1236	2	US-08-741-134-5
18	48.8	0.7	289	4	US-09-007-005-17
19	48.8	0.7	289	4	US-09-244-796-17
20	48.8	0.7	696	4	US-09-461-697-193
21	48.8	0.7	699	4	US-09-461-697-191
22	48.8	0.7	717	4	US-09-461-697-189
23	48.8	0.7	774	4	US-09-461-697-187
24	48.8	0.7	819	4	US-09-461-697-185
25	48.8	0.7	1669	4	US-09-461-697-184
26	48.6	0.7	1931	2	US-09-130-114-2
27	48.2	0.7	966	2	US-08-766-738-2

c 28	48.2	0.7	3211	2	US-08-574-959A-8	Sequence 8, Appli
c 29	48.2	0.7	3211	4	US-09-357-014-8	Sequence 8, Appli
c 30	48.2	0.7	3901	2	US-08-574-959A-6	Sequence 6, Appli
c 31	48.2	0.7	3901	4	US-09-357-014-6	Sequence 6, Appli
c 32	46.6	0.7	1052	1	US-08-466-603-1	Sequence 1, Appli
c 33	46.6	0.7	1052	1	US-08-314-503A-1	Sequence 1, Appli
c 34	46.6	0.7	1052	1	US-08-468-066-1	Sequence 1, Appli
c 35	46.6	0.7	1052	2	US-08-486-717-1	Sequence 1, Appli
c 36	46.6	0.7	1052	3	US-08-466-743-1	Sequence 1, Appli
c 37	46.6	0.7	1052	5	PCT-US95-12414-1	Sequence 1, Appli
c 38	46	0.7	18596	4	US-09-318-448-11	Sequence 11, Appli
c 39	44.8	0.6	9636	1	US-08-323-170B-1	Sequence 1, Appli
c 40	44.8	0.6	9636	4	US-08-954-441-1	Sequence 1, Appli
c 41	44.2	0.6	5163	3	US-08-700-651-1	Sequence 1, Appli
c 42	44.2	0.6	5163	3	US-08-928-361B-4	Sequence 4, Appli
c 43	44.2	0.6	5318	3	US-08-700-651-2	Sequence 2, Appli
c 44	44.2	0.6	5318	3	US-08-928-361B-3	Sequence 3, Appli
c 45	43.6	0.6	43795	3	US-08-742-185-101	Sequence 101, App

ALIGNMENTS

RESULT 1
US-08-616-844-7
; Sequence 7, Application US/08616844
; Patent No. 5849578
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A.
; TITLE OF INVENTION: COMPOSITION AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/616,844
; FILING DATE: 15-MAR-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,654
; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-053
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6407 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-08-616-844-7

Qy	1800	tcaatccttaacagagctctccacagagcaaacctctccagcccaacagcaccacttagc	1859
Db	2766	TCCATCTCTTAACAGAGTCTCCACAGAGCAAACTCTCCAGCCACAACTTAGC	2825
Qy	1860	acaaatgtctccaacttcacaaactaccattctgaagacctctgaagctcttatgaccac	1919
Db	2826	ACAAATGCTCTCGAACTTTCACAACATACCATTCTGAAGACCTCTAGCCTCTTATGACCAC	2885
Qy	1920	tcttgccacccctgtcaagcacagcatctctgtgctactgtgcccctatagcgttacagactac	1979
Db	2886	TCTTGGCACCCCTGTCAAGCACAGCATCTCTGTGTACTGTGCCTATAGCGGTACAGACTAC	2945
Qy	1980	agctggaaaaacagctctcgctgaccatactctgaaataactagttctctcaaatctcaacaga	2039
Db	2946	AGCTGGAAAAACGCTCTCGCTGACCCTCCTGAAATACTAGTTGCTTCAAAATCTCAACAGA	3005
Qy	2040	aggtggcatcagcacagaagaacagagtgatgtgtagctaccactggatgatccc	2099
Db	3006	AGGTGGGATCAGCACAGAAAGNACCGAGTGATTGTGGATGCTACCCTGGATTGATCCC	3065
Qy	2100	tttgaccagtgtagccacatcagcagcaaaagaaatgaccacaaagctttggcgttacagcaga	2159
Db	3066	TTTGACCAGTGTAACCCACATCAGCAAAAGAAATGACCACAAAGCTTGGCGTTACAGCAGA	3125
Qy	2160	gtacagccagcttcaagttccctcggaacatctctctctcccaaacacacagttgttct	2219
Db	3126	GTACAGCCAGCTTCAGTTCCTCTGGAACTCTCCTTCTCCCAAAACACAGTGTGTTTC	3185
Qy	2220	cacggtggaagcttggtctcccaaatctgcacactttgtctgcagagcagcacacagtc	2279
Db	3186	CACGGCTGAAGACTTGGCTCCCAATCTGCCACCTTTCGTGTTAGAGCAGCACACAGTC	3245
Qy	2280	accaacaacactgtcctctctcagctcagtcagcaacagctgtgtgtgaacctgtcttcca	2339
Db	3246	ACCAACAACACTGTCTCTCTTCAGCTCAGTCAACAGCTGTGCTGTGAACCTTGTCTTCA	3305
Qy	2340	caatgcggaatcgctcgacagacaacacagcgtggtaccactgcagggtcccgccttc	2399
Db	3306	CAATGGCGAATGGCTCGCAGACAACACAGCCGTGGCTTACCACATGCAGTGCCTGCGCTTC	3365
Qy	2400	ctggcaagggtatgttcagtgtagtgatgaatgcagtgctgtcgaaacctctgccatc	2459
Db	3366	CTGGCAAGGGGATGATTGCAGTGTGGATGTGAATCAGTGCCTGTGCAACCCCTGCCCATC	3425
Qy	2460	cacagccaogtgcacaa tactcagggtatcctttatctgcgcaaatgcocggttgggtacca	2519
Db	3426	CACAGCCACGTGCAACAATACTCAGGGATCCCTTTATCTGCAAAATGCCCGTGGGTACCA	3485
Qy	2520	gttggaaaaaggatagtaatttggttagaacctctgtgacagagtttaaataaagag	2579
Db	3486	GTTGGAAAAAGGATATGCAATTTGGTTAGAACCCTTCGTGACAGTTTAAATTTAAAGAG	3545
Qy	2580	aactttcttaatacactgtggaataacattcagacctcaagaagtgaataatgat	2639
Db	3546	AACTTTCTTAATACAACTGTGGAANAACATTCAGACCTACAGAAGTTGAAAATGAGAT	3605
Qy	2640	caccaaacgttaaatatgtgttttccagctgttaoctagttacatccgactacagttca	2699
Db	3606	CACCAAAACGTTAAATATGTGTTTTTCAGCGTTACCTAGTTACATCCGATCTACAGTTCA	3665
Qy	2700	cgctctagggagtcacacgcggttggtatctcactgcgaacaccttttccctggcctc	2759
Db	3666	CGCCTCTAGGGAGTCCAACGCGGTGGTGATCTCACTGCGAAACAACTTTTCCCTGGCCTC	3725
Qy	2760	caatgtacgctatttgacctggtcgtgataggatgcagaaatgtgtcaactcctgcaagtc	2819
Db	3726	CAATGTACGCTATTATGTGACCTGGCTGTATAGGATGCAGAAATGTGTCAACTCCTGCAAGTC	3785
Qy	2820	ctctgctgaggtctgcacgctcttggtatcttcagagcgaggtatcttagagcggcagctt	2879
Db	3786	CTCTGTGAGTCTGCCACGCTCTTGGGATCTCAGAGGCGGATCTTTAGACGGCGCACGTT	3845
Qy	2880	gtcgaagcggaagagtcgccgaatgtgacaagagacaccttccatctgcactgacctggacg	2939

[illegible]

Db	4926	AAACTGGCTGTGTTAAATATTTTCTGCGCCAGCRCTCGGTGCTAAACAACACAGATGAGGA	4988
Qy	4020	tgaagcgtaccactgaagctgaagatgtcgccattgaacgacagtgcttttcaatgcttt	4079
Db	4986	TGASCGTACCACGTGAAGTCTGAAGATGTGCGCATTTGAACGCACAGTGTTTTCATATGTTT	5045
Qy	4080	ctagggtgtcttatgtctacagtttccaagcagcccccaagtcagagaaatgtgtgaggc	4139
Db	5046	CTAGGTGTCTTATGTCTACAGTTTCCAAGCCASCCCACACAGTCAGGAAATGTGTGAGGC	5105
Qy	4140	accgcacacaactgcgaatgtgttttttaagtcaagtgacacatgatttaagatattttt	4199
Db	5106	ACCCACACAACTGCATGTGTTTAAAGTCACAGTCACACATGTATTAAAGATTTTTT	5165
Qy	4200	tttaaaatctcttgcagtttaaatctacttttcaaaagcctgatacagggcaaac	4259
Db	5166	TTTTAAAATCTCTTTCAGTTAAATCTCACTTTTCAAAACAAGCCTGGATCAGGGCAAAAC	5225
Qy	4260	aacttatattgttttaagctgagagctcagcagcagatgtcagcgagggggcaccttt	4319
Db	5226	AACTTATTGTGTTTACCTGGAGCTCAGCAGCGCATTTGCAGGACAGGGGGGCACTTT	5285
Qy	4320	tcatccatgagggccagcctggggcctgggaactctgataccacttgtgagggccagag	4379
Db	5286	TCATCCATGAGGGCCAGCCTGGGCGCTGGGACTCTGTATCACCATTGTGGAGGCCACAGG	5345
Qy	4380	cagctgcgtatgagagagaaaatgtcaaaactgaacgcaggtttccaccactctaggaagaca	4439
Db	5346	CAMCTGGGTATGAGAGAGAAATGTCAAACTGAACGCAGGTTTCACCACTCTAGGAAAGCA	5405
Qy	4440	gcttgtagccctcgcagctggaatgtggttagagggatgggctgaaataggcaggtttaga	4499
Db	5406	GCTGTGTTGACCCCTGCASCTGGATGTGTTAGAGGGATGGCTGAATAGSCAGGTTAGA	5465
Qy	4500	ttctctgatacaacagtgcttgggaagctgtgtggaatcctctgagggaagaacaggagcc	4559
Db	5466	TTTCTCGCATCAACAGTGTCTTTGGGAASCTGTGTGGATTCCTGAGGAAGAACACGGGAGCC	5525
Qy	4560	gagatggagccacacatgagttgtctcacgggctactgcagcacttgttaccagaactct	4619
Db	5526	GAGATGGAGCCACACATGAATTTGCTCACCGGCTACTCGACGACTTGTACCCAGAACTCT	5585
Qy	4620	catgtccacaaaccccatgtaaactttcaaccactcaagctgtttatctoggctgaagaa	4679
Db	5586	CATGTCCACAAACCCCATGTAAACTTCAACCACTCAAGSTGTTTATTTCGGCTGAAGAA	5645
Qy	4680	ataactttttctcacccagtcattgtactcttcatagtctgtgtcgacacctcc	4739
Db	5646	ATAACTTTTCTCACCCAGTCATTGTACTCTTCTATATGSGTATGTGCGCACCCCTCC	5705
Qy	4740	agaaacgtggtataacttccagtcagtgagggaactgaagactccggttgtgtagg	4799
Db	5706	AGAAACCTGGTTTACTKGCAGTCAAGTGTGGGAGAACTGAAGACTTCCGGTGTGTCGAGG	5765
Qy	4800	aactgaggggtgaccttcgggaaggaagttccactcatcttatattattatgcctgtgatg	4859
Db	5766	AACGTAGGGTTGACCTTCGGGAAGGAAGTTCACCACTCATCTTATTATTATGCGCTGTGATG	5825
Qy	4860	tgggtctgcagggagacatccagtaactcggtgtctttaatgtccactctggggaactgt	4919
Db	5826	TGGGTCTCTGCAGGGAGACATCCAGTACTCGGTGTCTKTAAATGCCACCTGCGGAACTGT	5885
Qy	4920	gtttattggcctcttttggggacatcctggttttagatgaagtggaggggaaatacagaggtta	4979
Db	5886	GTTTATTGTGGCTTCTTTTGGGGCATCTCGTGTTCGATGAAGTGAAGGGGAATACAGAGGTA	5945
Qy	4980	aagaaatgtctcacccctgaagcggggagtcgcccttccacttccattcttggaatggtcgag	5039
Db	5946	AAAGAATTGTCTCCACCTTGAAAGCGGGGAGTCCCCTCTCACATTTCTTGGAAANTGGTCAG	6005
Qy	5040	ccactggggagactctgcccggggcatggttgtttcttcaaggtctctctaaataataatc	5099
Db	6006	CCACTGGGACAGTTCTGCCCCGGGATCGTTGTTTCTTCAAGGTCTCTTAATATAATCT	6065

Qy	5100	cctattctacataactcttggccctgtatgggttttaagaagaagaactcctgtgtcccatgg	5159
Db	6066	CCTATTCTTACATAATCCCTGGCCCTGATGGTTTTTAAGCAAGAACTCCTGTGTCCTMAATGG	6125
Qy	5160	tctccaccactcaccatcacccctgctgtagcaagagtccttagtcagggaggtgcatttt	5219
Db	6126	TCGCCACCACTCACCATCACCCCTGCTGTAGCAAGAGTCCTAGTCAGGGGAGGTGCATTTT	6185
Qy	5220	agtagttaaatctgcacttatccatgagataataataaaaggagaagaactggttttatcagtgga	5279
Db	6186	AGTAGTTACATTCCACTTATCCATGAGATAAATAAAGAGAGAVCTGTTTTTATCAGTGGGA	6245
Qy	5280	ggcttaacctaaatttcaaaagtgtggccctttttgaaactcttgggacctctctctctgtaga	5339
Db	6246	GGCTAACCTTAATAATTTCAAAGTGTGCCTTTTTTGAATCTTTGGGCTCTCTCTCTGTAGA	6305
Qy	5340	acaaatggcccttctgtggtctcagcgccctcgacccctaaactgagagattctcgagctcctgca	5399
Db	6306	ACCAATGCCCTTTTGTGGCTCAGCGGCTCGCACCTTAACCTGGAGAGTTCTGAGCTCCCTGCA	6365
Qy	5400	gctcaactgagccacagactaggtctcttggctcctcttcgcg	5441
Db	6366	GCTCAGCTGAGCCACAGACTAGGCTTCTTGCTGCTCTCTTCGCG	6407

```

RESULT      2
US-08-599-654-7
; Sequence 7, Application US/08599654
; Patent No. 5882925
; GENERAL INFORMATION:
; APPLICANT:  FALB, DEAN A
; TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION:  TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES:  54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:  PENNIE & EDMONDS
; STREET:  1155 Avenue of the Americas
; CITY:  New York
; STATE:  New York
; COUNTRY:  USA
; ZIP:  10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE:  Floppy disk
; COMPUTER:  IBM PC compatible
; OPERATING SYSTEM:  PC-DOS/MS-DOS
; SOFTWARE:  PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER:  US/08/599,654
; FILING DATE:  09-FEB-1996
; CLASSIFICATION:  800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  US 08/485,573
; FILING DATE:  07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:  US 08/386,844
; FILING DATE:  10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME:  CORUZZI, LAURA A
; REGISTRATION NUMBER:  30,742
; REFERENCE/DOCKET NUMBER:  7853-041
; TELECOMMUNICATION INFORMATION:
; TELEPHONE:  (212) 790-9090
; TELEFAX:  (212) 869-8864
; TELEX:  66141 PENNIE
; INFORMATION FOR SEQ ID NO:  7:
; SEQUENCE CHARACTERISTICS:
; LENGTH:  6407 base pairs
; TYPE:  nucleic acid
; STRANDEDNESS:  both
; TOPOLOGY:  unknown
; MOLECULE TYPE:  CDNA
US-08-599-654-7

```


Query Match	73.3%	Score 5098.4	DB 2	Length 6407
Best Local Similarity	94.3%	Pred. No. 0		
Matches 5412	Conservative 21	Mismatches 8	Indels 301	Gaps 2
QY 1	gttcgataaagaattgcgcgtttttcaacaaaaagagtgaacagagcctcggagatgggaac	60		
DB 666	gttcgataaagaattgcgcgtttttcaacaaaaagagtgaacagagcctcggagatgggaac	725		
QY 61	agagagggcgatggggcgtgcagaagaatgactctgcacagcaagaggccaccacttc	120		
DB 726	agagagggcgatggggcgtgcagaagaatgactctgcacagcaagaggccaccacttc	785		
QY 121	ggcttggagcccttccttttctcctgttttgagatgggagctgacacagcctctctag	180		
DB 786	ggcttggagcccttccttttctcctgttttgagatgggagctgacacagcctctctag	845		
QY 181	gaagagaattccacagaccagatctcctcgtcgtatcttctacagagacagacttc	240		
DB 846	gaagagaattccctcagaccagatctcctcgtcgtatcttctacagagacagacttc	905		
QY 241	ctctcctcttagaccttctcctacctctgaaagtacagagaagcttaacaactccac	300		
DB 906	ctctcctcttctagaccttctcctacctctgaaagtacagagaagcttaacaactccac	965		
QY 301	tggcctccagagctcctcagtcagtcacaaagaacatgcattgtgtacacgtgtccac	360		
DB 966	tggcctccagagctcctcagtcagtcacaaagaacatgcattgtgtacacgtgtccac	1025		
QY 361	tgatgtggcccgagaaagctgcgacttttgacggtcagctcgtggacctctgagcaaac	420		
DB 1026	tgatgtggcccgagaaagctgcgacttttgacggtcagctcgtggacctctgagcaaac	1085		
QY 421	agaaggtctcccaaggactccagaattgccaagactcactcctcagtcctctcttcacc	480		
DB 1086	agaaggtctcccaaggactccagaattgccaagactcactcctcagtcctctcttcacc	1145		
QY 481	ctctgcagtggaatcgagaagaacagtagagtaactgggaatccaggggatggagaatt	540		
DB 1146	ctctgcagtggaatcgagaagaacagtagagtaactgggaatccaggggatggagaatt	1205		
QY 541	cattgaaacattccacagaaaaatgaattggacttaagctctttgag - tggcaaaatgattc	599		
DB 1206	cattgaaacattccacagaaaaatgaattggacttaagctctttgag - tggcaaaatgattc	1265		
QY 600	cccaaccttgagaacaacatcagcttgccagcagctctgagtgcaaaatggaagtcacat	659		
DB 1266	cccaaccttgagaacaacatcagcttgccagcagctctgagtgcaaaatggaagtcacat	1325		
QY 660	gtctcagactgagactgtctctagtcagtcgacccatgagagtgagagatcaactgc	719		
DB 1326	gtctcagactgagactgtctctagtcagtcgacccatgagagtgagagatcaactgc	1385		
QY 720	acaatggctcttgaccaaacagcaacaactcgcagatgtgacaggaagctcgtctcata	779		
DB 1386	acaatggctcttgaccaaacagcaacaactcgcagatgtgacaggaagctcgtctcata	1445		
QY 780	tccctgaagtgatgaactcagctgagtgaccaggtctcagacactcactgctacagctcgg	839		
DB 1446	tccctgaagtgatgaactcagctgagtgaccaggtctcagacactcactgctacagctcgg	1505		
QY 840	aggaagtccacagcattgggagatagaggtattcagagctcttcacatctactctctc	899		
DB 1506	aggaagtccacagcattgggagatagaggtattcagagctcttcacatctactctctc	1565		
QY 900	ggaagcttgaaattcatcagcaccacgctggagaacgcttcaa - - - - - 940			
DB 1566	ggaagcttgaaattcatcagcaccacgctggagaacgcttcaa - - - - - 940	1625		
QY 941	----- 940			
DB 1626	gccaggccaaagcactagtgtagacttcgcgccaattcgagagcagaggacttctgggggtgcc	1685		

QY 1800 tcaatccttaacagagtcctccacagacgaacaccccttcagccacacgaacccaattgac 1859
Db 2766 TCCATCCTTAACAGAGTCCTCCACAGAGCAAAACCCCTCCAGCCACAAGACCAACTAGC 2825
QY 1860 acaaatgtctccaaacttcaaaactaccattctgaagaacctctcagcctcttatgaccac 1919
Db 2826 ACAAAATGCTCCAAACTTTTCAAACTACCAATTCATGGAAGACCTCTCAGCCCTTTATGACCAC 2885
QY 1920 tcttgccacctgtcaagcacagcatctctgttcaactggccctctatagccgtacagactac 1979
Db 2886 TCCTGGCACCTGTCAAGCACACACATCTCTGGTCACCTGGCCCTATAGCCGTACAGACTAC 2945
QY 1980 agtggaaaacagctctcgtgacccatcctgtaaa tactagttctcctcaaatctcaacaga 2039
Db 2946 AGGTGGAANAACAGCTCTCGTGTACCCATCCTGAAATACTAGTTCCCTCAAAATCTCAACAGA 3005
QY 2040 agtggcatcagcacagaagaacagcagtgattgtggatgctaccacitggatgtatccc 2099
Db 3006 AGGTGGCATCAGCACAGAAGGAACCGAGTGATTGTGGATGCTTACCACATGGATTGATCCC 3065
QY 2100 ttgacccagtgtaaccacatacagcaaaagaaaatgacccaaaagccttggcgttacagcaga 2159
Db 3066 TTTGACCAAGTGTACCCACATCAGCAAAAGAAATGACCAAAAGCTTGGCGTTACAGCAGA 3125
QY 2160 gtacagccagcttcacgttccctcgtggaacatctcctctcccaaacacacagttgtttc 2219
Db 3126 GTACGCCAGCTTCACGTTCCCTCGAATCTCCTTCTCCCAAAACCAACACAGATTGTTTC 3185
QY 2220 cagcgtgaagacttgctcccaaaactcaccaccttggctgttcagagcagacacagtc 2279
Db 3186 CAGCGCTGAAGACTTGGCTCCCAAACTCGCCACCTTTCCTGCTTTCAGAGCAGCACAGTC 3345
QY 2280 accaaacaaactgtcctcttcagcctcagtcacacagctgtgctgtgaaccttggcttcca 2339
Db 3246 ACCAAACAACACTGCTCTTTCAGCCCTCAGTCAACAGCTGTGCTGTGAACCCCTTGTCTTCA 3305
QY 2340 caatggcgaatggcgtcgagacaaacccagcgcgtgctaccactgaggtgcgcgccttc 2399
Db 3306 CAATGGCGAATGGCTCGCAGACAAACACAGCCGTGCTTACCAGTGAAGGTGCCCCGCTTC 3365
QY 2400 ctggcagggagtgattgcaagtggatgtgaatgagtgctgcgaacccctgcgcacac 2459
Db 3366 CTGGCAAGGGGATGATGCAAGTGGATGTGAATGAGTGCCTGTCCGACCCCTGCCCTNYC 3425
QY 2460 cacagccaogtgcacaa tactcagggatccttatactgtcaaaatgccccgggtgggtacca 2519
Db 3426 CACAGCCACGTGCAACAATACTCAGGGATCCTTTATCTGCAAAATGCCCGTTGGGTACCA 3485
QY 2520 gttggaaaaaggatgatgaatttggtagaaccttcgtgacagagtttaaaattaaagag 2579
Db 3486 GTTGGAAAAAGGGATATGCAAAATTTGGTTTGAACCTTCGTGACAGAGTTTAAATTAAGAG 3545
QY 2580 aactttcttaaTacaactgtgaaaaacattcagacctacagaagattgaaaatgagat 2639
Db 3546 AACTTTTCTTAATACAACTGTGGAANAACATTCAGACCTTACAAAGATTGAAAAATGAGAT 3605
QY 2640 caccaaaacgttaaatatgtgttttccagcgttacctagttacatccgatctcacagtcca 2699
Db 3606 CACCAAAACGTTAAATATGTGTTTTTTCAGCGTTACCTAGTTACATCCGATCTACAGTTCA 3665
QY 2700 cgctctagggagtcacaagcgggtggtgatctcactgcaaaacacccctttccctcgccctc 2759
Db 3666 CGCCTCTAGGGAGTCCAACCGGTGGTGTATCTCAGTGCAAAACAACCTTTTCCCTGGCCCTC 3725
QY 2760 caatgacagctatttgacctgctataggaatgcagaaatgtgtcaactcctgcgaagtc 2819
Db 3726 CAATGTGACGCTATTGACCTGGCTGATAGGATGCAGAAATGTGTCAACFTCCCTGCAAGTC 3785
QY 2820 ctctgctgaggtctgcagctctctggatctcagagcggcgatcttttagacggcgagcgtt 2879
Db 3786 CTCTGCTGAGGTCTGCCAGCTCTTGGGATCTCAGAGCGGATCTTTAGAGCGGGCAGCTT 3845
QY 2880 gtgcaagcggaaagatcccgaaatgtgacaaagacacctccatctgcactgacctggacgg 2939

Db 3846 GTCAACGCGAAAGAGTCCGAAATGTGACAAAGACACTTCCATCTGCACCTGACCTGGACGG 3905
QY 2940 cgttgcctcgtgcagtcgaagtcggatacttctcagttcaacaagatgaccactcctg 2999
Db 3906 CGTGTGCCCTGTGCCAGTGCNAAGTCGGGATACTTTTCTAGTTCAACAAGATGACCACTCCTG 3965
QY 3000 ccgagcatgtgaagatggatataggtctgaaatgaaacctgcattgattggccatttgg 3059
Db 3966 CCGAGCATGTGAAGATGGATATAGGCTTGAATGAACCTGATGATGAGTTGCCCATTTGG 4025
QY 3060 ccttggtgctcaactgtggaaaacccctatacagcttatacctgtggtgatacgcgcgc 3119
Db 4026 CCTTGTGTGTCTCAACTGTGAAACCCCTATCAGCTTATCAGCTGTGTGTATCCGACGCCG 4085
QY 3120 gggagtggtgctcctcctcatcctcctagcctgcacgtgattgttacctgttgccagaagaa 3179
Db 4086 GGGAGTGGGCTCCTCTCTCTCTAGCATCGCACATGATTGTTACCTGTTGAGANAAGAA 4145
QY 3180 taaaaatgacataagcaaaactcatttcaaaagtggagatttcccaatgttccccatgc 3239
Db 4146 TAAAAATGACATTAAGCAAACTCATCTTCAAAAGTGGAGATTTCCAAATGTCCCATATGC 4205
QY 3240 tgaatcccccaaaaaatcctcgtcacaagaatggggcccgagaagctattgaaatgcatga 3299
Db 4206 TGAATACCCCAAAAAATCCTCGCTCACAAAGATGGGCGGAGAGCTATTGAAATGCAATGA 4265
QY 3300 gaatgaagtaccaaaaacccctcctccagatgcaggtatgtgtactactcgcctcacaagtgt 3359
Db 4266 GAATGAAGTACCAAAAACCTCTCCAGATGACGGATGTGTACTACTCGCCTACAGTGT 4325
QY 3360 aaggaatccagaacttgaacgaacggaactctaccgggctacactggagctgcaggaatc 3419
Db 4326 AAGGAATCCAGAATTTGAACGAACGAGCTCTACCCGGCTTACACTGGAGCTGCAGGATC 4385
QY 3420 acggcattcttgatcttcccgagacagtataaacctgttctcactcagtgatgaaagcag 3479
Db 4386 ACGGCATCTGTGATTTTCCCGGACAGTATAACCCGCTCTTTCATCAGTATGAAGCAG 4445
QY 3480 aagaagagactactttaaagtcacagagagagaggaactcattgctcgtgagccagtcacc 3539
Db 4446 AACAAGAGACTACTTTTAAGTCCAGGAGAGAGAGGACTCATTTGCTGTGAGCAGTCAAC 4505
QY 3540 tgggacctctgtcagaggaacccacagagagctgcgcccagattgttcgggagccac 3599
Db 4506 TGGGACCTCTGCTCAGAGGACCCACAGAGGCTGCGCCGAGATTTGCGGGAGCCAC 4565
QY 3600 gctgagtggaagcaggaagagagcagcagcagcagcagcagcagcagcagcagcagcagc 3659
Db 4566 GCTGAGTGGCAACGAGAGAGGACAGGATGCGGGGCGGTGACCACTGTGAGGAGACA 4625
QY 3660 ggtggatgtggaaccacagcgtcctcattcagcacctttgttactgctgaaactgaaat 3719
Db 4626 GGTGGATGTGGAACACACAGCTGCTCATTCAGCACCTTTTGTGTTACTGTGAACGTGAAT 4685
QY 3720 gtggcgcagtatcaagagagctctctctgactgacccatggcaccatggcaccagggcg 3779
Db 4686 GTGGCCAGTATCAAGAGAGTCTCTCTGAGTACTGCACCATGGCATGGCAGGCGG 4745
QY 3780 actattagcagggcagaccactagacttcaagtcagggagacctggtttccctcgtttg 3839
Db 4746 ACTATTAGCCAGGGCAGACCACTAGACTTCAGTGCAGGGACCTGGTTTCCCTTCGTTG 4805
QY 3840 cacttagtaaaattgggtggaggtttcccttttggatcctgttttgagactgtttccagaaa 3899
Db 4806 CACTTTAGTAAATGGTGGGAGGTTTCCCTTTGGATCTCTTTTGAGACTGTGTCCAGAAA 4865
QY 3900 gaaggttctcttcccgagacacttccatagggcagcaaatgttgatatttcatttgcagcaa 3959
Db 4866 GAAGGCTTCTTTCCCGAGACACTTCCATAGGCAGCAATTTGGTGAATTCATTTTGCASAA 4925
QY 3960 aaactggtgtgtaattatttctgcccagcgcctgctgcttaaacacacagatgagga 4019

Db 4926 AATAGTGGCTTGTAAATTAATTTCTCTGCCAGRCCTGCGTGCTAAACAACAGATGAGGA 4985
Qy 4020 tgacgtaccactgaagctgaagatctgccattgaacgagacagtgttttcatatattt 4079
Db 4986 TGACGTAACCACTGAAGTCTGAAGATGCGCATGTAACGACAGTGTCTTTCATATGTTT 5045
Qy 4080 ctaggtgtctttagctacagtttcccaagcagcccccacagtgaggaatgtgtgagcc 4139
Db 5046 CTAGGTTGCTTATGCTACAGTTTCCAAGCCASCCCCACAGTGAGGAATGTGTGAGGC 5105
Qy 4140 accgcacacactgcaatgtgtttttaaagtcgaaggtgacacatgtatttaagattttt 4199
Db 5106 ACCGCACACAACTGCAATGTGTCTTAAAGTCAAGGTGACACATGTATTAAGATTTTTT 5165
Qy 4200 ttttaaatctcttgagtttaaatctcaccttttttcaacaagcctggatcagggcacaaac 4259
Db 5166 TTTAAATCTCTTTCAGTCTAAATCTCACTTTTCAAAACAAAGCCTGGATCAGGCGCAAAAC 5225
Qy 4260 aacttatatttggttttagctggaggtctcagcagcgagatgcaagcagggggcactttt 4319
Db 5226 AACTTATATYTGTTTAGCTGGAGGCTCAGCAGCGAGATGCGAGCGGGGGCACATTT 5285
Qy 4320 tcattcatgagggcccagcctggggcctgggaactctgatcaccaattgtgagggccagag 4379
Db 5286 TCATCCATGAGGGCCAGCCTGGGCGCTGGGACTCTGATCACCATTTGTGGAGGCCAGAGG 5345
Qy 4380 cagctgcatgagggagagaatgtcaaacgtgaacgaggtttcaaccactctagaaagca 4439
Db 5346 CAMCTGCGTATGGAGGAGAAATGTCAAACTGAACGCAAGTTTCAACCACTCTAGAAAGCA 5405
Qy 4440 gcttgtagccctcagctggatgtggttagaggtggctgaatagcaggttaga 4499
Db 5406 GCTTGTTGACCCCTGCASCTGGATGTGTTAGAGGATGGGCTGAATAGSCAGGTTAGA 5465
Qy 4500 ttctctgcatcaacagctgctttgggaagctgtgtggattcctgaggaagaacagggagcc 4559
Db 5466 TTTCTCTGATCAACAGTCTTTGGGAASCCTGTGTGGATTCTCTGAGGAAGAACAGGAGCC 5525
Qy 4560 gagatgagccacacatgattgtctcaccggctactgcagcactttgtaccccaagaatct 4619
Db 5526 GAGATGAGGCCACACATGAATGTGCTCACCCTGCTACCTGCTGCTGCTGCTGCTGCTGCT 5585
Qy 4620 catgtccacaaacccctgtaaaccttcaaccactcaaacctgttttctggctgaagaa 4679
Db 5586 CATGTCACAAACCCCATGTAAACTTCAACCACTCAAAAGTGTATTCGGCTGAAGAA 5645
Qy 4680 ataacttttttctcaccagctattgtacctcttctcatatggtctgtctgcaccctcc 4739
Db 5646 ATAACCTTTTCTCACCAGTCAATTTGTACCTCTTCAATATGCTGCTGCTGCTGCTGCT 5705
Qy 4740 agaaactgtgtatacttcacgtcagtggtgggaagactgaagactccggttggtcaggg 4799
Db 5706 AGAAACGTGGTTATACATKCCAGTCAGTGTGGGGAAGTGAAGACTCCCGTGTGCTCAGG 5765
Qy 4800 aactgaggggtgaccttcgggaagagttccactctctatttatttattgctgtgatg 4859
Db 5766 AACTGAGGGTTGACTTCGGGAAGAGTTCACCTATCTTATTTATTTATGCTGCTGATG 5825
Qy 4860 tgggtcctgcagggagacatccagtagtctcgtgtgttatttattgcccactcgggaaactgt 4919
Db 5826 TGGGTCTCTGCGAGGAGACATCCAGTACTCGGTCTCTAATTTCCACCTCGGGGAACCTGT 5885
Qy 4920 gtttattggcctcttttggggcatcctgggttttgatgaagtgaggggaatacagaagta 4979
Db 5886 GTTTATTGGGCTCTTTTGGGGCATCTCTGGKTTGCGATGAAGTGAAGGGAATACAGAGGTA 5945
Qy 4980 aaagaattgtctcacctgaagcggggagtcctcgcttccacttctggaatggtcag 5039
Db 5946 AAAGAATTGTCTCACCTGAGGGGGAGTCCCGCTTCACATTTCTGGAAATGCTGCAG 6005
Qy 5040 ccactggggagcttctgccccgggagctgtgttttcttcaaggtctctataatc 5099
Db 6006 CCACTGGGACAGTCTGCCCCGGGCATGTTGTTTCTTCAAGGTCTCTAAATATAATC 6065

RESULT 3

US-08-944-868A-7
; Sequence 7, Application US/08944868A
; Patent No. 6018025
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,868A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/599,654
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-041
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6407 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-08-944-868A-7

Query Match										73.3%										Score 5098.4;										DB 3;										Length 6407;																																																																																															
Best Local Similarity										94.3%										Pred. No. 0;										Matches 5412;										Conservative										21;										Mismatches										8;										Indels										301;										Gaps																																			
Qy	1	gttcgatgaagaattccgccttttccaaacaagaagctggaacagctctcgagatctgggaac	60	Db	666	CTTCGATGAAGAATTCGCCGCTTTTCAACAAGAAGTGAACAGCCTCGGAGATGGGAAC	725	Qy	61	agagaggcgatggggctgcagaagaatggactgtgcacagccaagagggccaccacttc	120	Db	736	AGAGAGGCGATGGGCTGTCAAGAAATGGACTGTGCACAGCAAGAGGCCACCACATTC	785	Qy	121	ggcttgagccctctcttttctctctctgtttggagatgggagagctgacacagcctctctag	180	Db	786	GGCTTGGAGCCGTCCTTTCTTCTGCTTGGAGATGGGAGAGCTGACACGCCCTTCTAG	845	Qy	181	gaagagaaatctctcaggaaccagatctctctggctgctattctacagagcagcagcttc	240	Db	846	GAAGAGAAATTCCTCAGGACCAGATCTCTCTGGCTTGCATTTCTACAGGACAGCAGCTTC	905	Qy	241	ctctctctcttagaaccttctctcaaccttctgaagctacagagaagcttaacaaactccac	300	Db	906	CTCTCCTCTTATGACTTTCTCCTCACTTCTGAAAGTACAGAGAAGCTTAAACAAGTCCAC	965	Qy	301	tggcctcagagctcctcagtcagtcagtcacaaagaacaaatgcagtgtgtacctgtttcac	360	Db	966	TGGCCTCCAGAGCTCTCAGTCACTAGTCAACAAAGACAAATGCATGTTGCTTACCGTGTTCAC	1020	Qy	361	tgatgtggcccgagaacgctgcgactcttgacggctcagctctgggacctgtgagcaagac	420	Db	1026	TGATGTGGCCGAGAAAGCTGCGCATCTTTCAGCGTCACTCTGGGACCTGTGAGCAAGAC	1080	Qy	421	agaagctctcccaaggactccagaattgcccgacctctatcctcagtcctctctctacc	480	Db	1086	AGAAGGCTTCCCAGGACTTCCAGAAATGGCCAGACTTCATCTCAGTCCCTTCTTTTCACC	1140	Qy	481	ctctgcagtggaattcgagaagaacagtagagtaactgggaaatccaggggagtgaggaatt	540	Db	1146	CTCTGCAGTGGAAATCGAGAGAAGAACAGTAGAGTAACCTGGGAATCCAGGGGATGAGGAAT	1200	Qy	541	cattgaaccatccacagaaaaatgaaattggacttaogtcttgctg-tggcaaaatgattc	599	Db	1206	CATTGAACCATCCACAGAAATGAATTTGGACTTAGTCTTTTGGCTTGGCAAAATGATTC	1265	Qy	600	cccaacctttggagaacatcagcttgcgcagcagctctgaggctgcaaaatggaagtcacct	659	Db	1266	CCCAACCTTTGGAGACATCAGCTTCCGACGAGCTCTGAGGTTGCAAAATGGAAGTCCCAT	1329	Qy	660	gtctcagactgagactgtgtctaggctcagtcgacccactgagaggtgagagatcactgctc	719	Db	1326	GTCTCAGACTGAGACTGTGTCTTAGTTCAGTCGACCCCATGAGAGGTGGAGAGATCAGTGC	1389	Qy	720	acactggctcttgaccaacagacaacactctcagatgtgacaggaagctctgctcata	779	Db	1386	ACACTGGCTCTTGACCAACAGCAACAATCTGCAGATGTGACAGGAAGCTCTGCTTTCATA	1445	Qy	780	tcctgaaggtgtaattgcttcaggttgaccagttctcagactctactgtacagctctgg	839	Db	1446	TCCTGAAGGTGTGAATGCTTCACTGCTTACCAGATTCACAGACTCTACTGTACAGCTCTGG	1505	Qy	840	aggaagtcaacagcattggggagataggagttattcagagttctcatcatctctctc	899	Db	1506	AGGAAGTTCACACAGCAATGGGAGATAGGAGTATTTTCAGAGCTTCTCATCTACATCTCTC	1565	Qy	900	ggaagcttgaattcatcagcaccacgctggagaaagcttcaa	940	Db	1566	GGAAAGCTTGAAATTCATCAGCACCAGCTGGAGAAAGCTCAACCTTGGGAAGACAGCCGAGA	1625	Qy	941	-----	940	Db	1626	GCCAGGCCAAGCACTAGGTGACAGTTCCGCCCAATGCAGAGGACAGGACTTCTGGGTGSCC	1685

Qy	941	-----	-----	-----	940
Db	1686	CTCTCTGGCACCCACACACCTTGGCTACTGTCACTGAAACGGGAACGCACACTGCGGTC			1745
Qy	941	-----	-----	-----	940
Db	1746	TGTCACCTCACC AACACACAGCATGACAGGACTTCTGGGAAGCAGGAGCCCTGCAAC			1805
Qy	941	-----	-----	-----	940
Db	1806	GGCCATGCCCAAGAAACAGAGGGTGCTCTCTGCACGTAAACGTACGGACGACATGGG			1865
Qy	941	-----	-----	-----	959
Db	1866	CCTGGTCTCAGGTCACCTGGCCGCTCCACGTGCACCTGGAGTGGCTGGGATTAGCTACGG		-----TCGCTGGGATTAGTCAcgg	1925
Qy	960	tcaagtgcgtgcacagctattgaacaaaggacttccagcagcaccacagaccacaccta			1019
Db	1926	TCAAAGTGGTGGCACACGCTATTGAACAAAGGACTTCCACGACACACAGACCACACCTA			1985
Qy	1020	cctgtcatctactttcaccaaaaggagaagcggcggttactgtccattacagataaacagttc			1079
Db	1986	CCTGTCATCTACTTTCACCAAGGAAGACGGCGGTACTGTCCATACAGATAACAGTTC			2045
Qy	1080	atccctcagacattgtgagagctcaactcttatataaaattctaaactctcaactctcaactc			1139
Db	2046	ATCCTCAGACATTGTGGAGAGCTCAACTTCTATATTAANAATCTCAAACTCTTCACATTC			2105
Qy	1140	agagattctctctcttctcatgctcagactgagagaagtaacatctcatctatgacgg			1199
Db	2106	AGAGTATTCCTCCTTTTCTCATGCTCAGACTTGAGAGAAGTAACATCTCATCTATGACGG			2165
Qy	1200	ggaatagctcagccttctactagtcgcgaagtctgcatacatcaacacctccgctcta			1259
Db	2166	GGANATGTCTCAGCCTTCTACTGAGTCGCCAGTTCTGCATACATCCAACTTCCGTCCTA			2225
Qy	1260	cacaccacattaatatgcgaacacttcggtgtgtcttgacactgagtctgagttgt			1319
Db	2226	CACACCCACCATTAATATGCGGAACACTTCGGTGTGTTCTGGACACTGATGCTGAGTTGT			2285
Qy	1320	tagtact			1379
Db	2286	TAGTGACT			2345
Qy	1380	gcctctgcctctgtgtcacaatcccaaccattattttcacaatttaccatacaacag			1439
Db	2346	GCCCTCGCCCTGTGTFCAAATCCCAACCATTAATTTTCATCAATTTTACCATCAACAG			2405
Qy	1440	ggcctctgtgcatactactaaagtctacctctgtatgcatacacacacatggttctctcaac			1499
Db	2406	GGCCTCTGTGCATCTACTAAGTCTACCTCTGATGCATCCACACCATGGTCTCTCTCACC			2465
Qy	1500	atacctttaccagtatccttaacgacatctacatctgcccaacttctgtctctcacaac			1559
Db	2466	ATCACCTTTACAGTAGTCCCTTAACGACATCTACATCTGCCCCACCTTCTGTCTCAAAAC			2525
Qy	1560	aaccttgccacagtcatcttctaccctgtctctgccagggaaggagactcctgtgac			1619
Db	2526	AACCTTGCCACAGTCATCTTCTACCCCTGTCTGCCAGGGCAAGGGAGACTCTCTGTGAC			2585
Qy	1620	tteatttcagaacatacaaatgacatcatctatgcacaatgtctccatagtagtcaaaatgc			1679
Db	2586	TTCAATTTACAGCATACAAATGACATCAPTTCATGACAATGCTCCATAGTAGTCAAACTGC			2645
Qy	1680	agaccttaagagccagagcaccaccaaccaagagaaagtcatctacagaaaccaagatcaac			1739
Db	2646	AGACCTTAAGAGCCAGACGACCCACACCAAGAGAAAGTCAATACAGAAATCAAAAGTCAAC			2705
Qy	1740	agacctggtgtctctgtccacagagtccaaccaaaagtctlaacaaactctctcttgc			1799
Db	2706	AAGCCTGGTGTCTCTGCCACAGTGCACAAAGCTGAACAAAGCTGAACAAAGCTCTCTCTTGC			2765

Db 4926 AATACTGGCTGTGTAATATATTTCTCCAGCGRCTCGCTGCTAAACAACATGAGGA 4985
QY 4020 TGAGCTGACCACTGAAGTCTGAAGATGCTGCCATTGAACGAGCAGTGTCTTCCATATGTTT 4079
Db 4986 TGASCGTACCACGAACTCTGAAGATGCTGCCATTGAACGAGCAGTGTCTTCCATATGTTT 5045
QY 4080 CTAGGTTGTTCTTATGCTACAGTCTTCAAGCCAGCCGCCACAGTCAGGAATGTGTAGGC 4139
Db 5046 CTAGGTTGTTCTTATGCTACAGTCTTCAAGCCAGCCGCCACAGTCAGGAATGTGTAGGC 5105
QY 4140 ACAGCACAACAGTCAAGTCTTCAAGCCAGCCGCCACAGTCAGGAATGTGTAGGC 4199
Db 5106 ACCGCACACAACTGCAATGTGTTTAAAGTCAAGGTGACACATGTATTTAAAGATTTT 5165
QY 4200 TTAAGATCTCTGCAAGTCTTCAAGCCAGCCGCCACAGTCAGGAATGTGTAGGC 4259
Db 5166 TTTAAATCTCTGCAAGTCTTCAAGCCAGCCGCCACAGTCAGGAATGTGTAGGC 5225
QY 4260 AACTTATATGTTTGTAGCTGAGGCTCAGCAGGAGATGTCAGCAGGGGGGACTTT 4319
Db 5226 AACTTATATGTTTGTAGCTGAGGCTCAGCAGGAGATGTCAGCAGGGGGGACTTT 5285
QY 4320 TCAATGATGAGGCGCCAGCCTGGGGCTGAGGCTCTGATCACCATTGTGAGGCCAGAGG 4379
Db 5286 TCAATGATGAGGCGCCAGCCTGGGGCTGAGGCTCTGATCACCATTGTGAGGCCAGAGG 5345
QY 4380 CAGCTGCTATGAGGAGAAATCAAACTGAACGAGGTTCCACCTAGGAAAGCA 4439
Db 5346 CAGCTGCTATGAGGAGAAATCAAACTGAACGAGGTTCCACCTAGGAAAGCA 5405
QY 4440 GCTTGTGAGCCCTGCASCATGATGTTAGAGGATGGCTGAATAGCAGGTTAGA 4499
Db 5406 GCTTGTGAGCCCTGCASCATGATGTTAGAGGATGGCTGAATAGCAGGTTAGA 5465
QY 4500 TTTCTGATCAACAGTCTTGGGAAGTCTGTTGAGGATGCTGAGGAGAACAGGGAGCC 4559
Db 5466 TTTCTGATCAACAGTCTTGGGAAGTCTGTTGAGGATGCTGAGGAGAACAGGGAGCC 5525
QY 4560 GAGTGGAGCCACATGATGTTGCTCAGCGCTACTGAGCAGCTTGTACCCAGAACTCT 4619
Db 5526 GAGTGGAGCCACATGATGTTGCTCAGCGCTACTGAGCAGCTTGTACCCAGAACTCT 5585
QY 4620 CATGTCACAAACCCCATGTAATTTCAACACTCAAGSTGTTATTTCGGCTGAAGAA 4679
Db 5586 CATGTCACAAACCCCATGTAATTTCAACACTCAAGSTGTTATTTCGGCTGAAGAA 5645
QY 4680 ATAAGTCTTCTTCAACAGTCTTGGGAAGTCTGTTGAGGATGCTGAGGAGAACAGGGAGCC 4739
Db 5646 ATAAGTCTTCTTCAACAGTCTTGGGAAGTCTGTTGAGGATGCTGAGGAGAACAGGGAGCC 5705
QY 4740 AGAAGCTGTTTACCTCCAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAG 4799
Db 5706 AGAAGCTGTTTACCTCCAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAG 5765
QY 4800 AACTGAGGTTGACCTTCGGAAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAG 4859
Db 5766 AACTGAGGTTGACCTTCGGAAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAG 5825
QY 4860 TGGGCTCTGCGAGGAGACATCCAGTACTCGGTGCTCTGAGGAGTCTGAGGAGTCTGAGGAG 4919
Db 5826 TGGGCTCTGCGAGGAGACATCCAGTACTCGGTGCTCTGAGGAGTCTGAGGAGTCTGAGGAG 5885
QY 4920 GTTATGCGCTCTTCTGCGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAG 4979
Db 5886 GTTATGCGCTCTTCTGCGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAG 5945
QY 4980 AAGAAATGCTCCACCTCAAGCGGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAG 5039
Db 5946 AAGAAATGCTCCACCTCAAGCGGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAG 6005
QY 5040 CCACTGGGAGTCTGCGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAG 5099
Db 6006 CCACTGGGAGTCTGCGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAG

QY 5100 CCTATTCTTACATAAATCTTGGCCCTGATGGTTTAAAGCAAGAACTCTGTGTCCTCCATGG 5159
Db 6066 CCTATTCTTACATAAATCTTGGCCCTGATGGTTTAAAGCAAGAACTCTGTGTCCTCCATGG 6125
QY 5160 TCTCCACCACTCAACATCAGCTGCTAGCAAGAGTCTTAGTCAAGAGTCTTAGTCAAGAGTCTTAGT 5219
Db 6126 TCTCCACCACTCAACATCAGCTGCTAGCAAGAGTCTTAGTCAAGAGTCTTAGTCAAGAGTCTTAGT 6185
QY 5220 AGTAGTTAAATGCACTTATCCATGAGATAAATAAAGAGAGTCTTAGTCAAGAGTCTTAGTCAAGAGT 5279
Db 6186 AGTAGTTAAATGCACTTATCCATGAGATAAATAAAGAGAGTCTTAGTCAAGAGTCTTAGTCAAGAGT 6245
QY 5280 GGTCAACCTAAATTTCAAGTGTGCTTCTTGAAGTCTTGAAGTCTTGAAGTCTTGAAGTCTTGAAGT 5339
Db 6246 GGTCAACCTAAATTTCAAGTGTGCTTCTTGAAGTCTTGAAGTCTTGAAGTCTTGAAGTCTTGAAGT 6305
QY 5340 ACCAATGCCCTTTGTGCTCAGCGCTCGACCTAACTGAGAGTCTTGAAGTCTTGAAGTCTTGAAGT 5399
Db 6306 ACCAATGCCCTTTGTGCTCAGCGCTCGACCTAACTGAGAGTCTTGAAGTCTTGAAGTCTTGAAGT 6365
QY 5400 GCTCACTGAGCCACAGTCTTCTTGGCTCTTCTTGGCTCTTCTTGGCTCTTCTTGGCTCTTCTTGGCT 5441
Db 6366 GCTCACTGAGCCACAGTCTTCTTGGCTCTTCTTGGCTCTTCTTGGCTCTTCTTGGCTCTTCTTGGCT 6407

RESULT 4
US-08-944-423A-7
; Sequence 7, Application US/08944423A
; Patent No. 6020463
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,423A
; FILING DATE: 06-OCT-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/599,654
; FILING DATE: 09-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: JUN-07-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CORUZZI, LAURA A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-105
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELE: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6407 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both

;
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
US-08-944-423A-7

Query Match 73.3%; Score 5098.4; DB 3; Length 6407;
Best Local Similarity 94.3%; Pred. No. 0;
Matches 5412; Conservative 21; Mismatches 8; Indels 301; Gaps 2;

QY	1	gttcgatgaagaattgcgccttttcaacaagaagatggaacggcctcgagatgggaac	60	QY	941	-----	940
Db	666	gttcgatgaagaattgccccttttcaacaagaagatggaacggcctcgagatgggaac	725	Db	1626	gccagcccaagacactaggtgacagattccggccaatgcagagacagagacttcttgggggtgcc	1685
QY	61	agagagggcgatgggctgacagaagaatgactgtgcacagccaagagggccaccattc	120	QY	941	-----	940
Db	726	agagagggcgatgggctgtgcagagnatggactgtgcacagccaagagggccaccattc	795	Db	1686	ctctctcgggcaccacacacacttggctactgtcacttggaaacggggaacgcacacacttgcggtc	1745
QY	121	ggcttgagcccttccttcttctctctgttggagatgggagactgaccacgcctcttag	180	QY	941	-----	940
Db	786	ggcttgagcccttccttcttctctctgttggagatgggagactgaccacgcctcttag	845	Db	1746	tgtcacccctcaccaacaccagcatgagcacgacttcttggggaacggagcagccctgcagc	1805
QY	181	gaagagaaattctcaggaaccagatctctctgctgctgattctacagagacagacttc	240	QY	941	-----	940
Db	846	gaagagaaattctcaggaaccagatctctctgctgctgattctacagagacagacttc	905	Db	1806	ggccatgcccccaagaaacagaggggtgcctctctgcacgttaaacgtgacggacgacatggg	1865
QY	241	ctctctctcttagaccttctcactctctgaagtcacagagaagcttaacaactccac	300	QY	941	-----	959
Db	906	ctctctctcttagaccttctcactctctgaagtcacagagaagcttaacaactccac	965	Db	1866	cctgggtctcactgcgtcacttggccgcctccagtcgactcgggagtcgctgggattagctacgg	1925
QY	301	tggcctccagagctctcagtcagtcacaaagaacaatgcatgttgcctaccgctttcac	360	QY	960	tcaagtgcgtggcacagctattgaacaaagacttccagcgacccacacagacacaccta	1019
Db	966	tggcctccagagctctcagtcagtcacaaagaacaatgcatgttgcctaccgctttcac	1025	Db	1926	tcaagtgcgtggcacagctattgaacaaagacttccagcgacccacacagacacaccta	1985
QY	361	tgatggtgcccagagaacgctggatctttgacggctcagtcctgggacctgtgagcaagac	420	QY	1020	cctgtcatctacttccacaaagagaacggggttactgtcattacagataacagttc	1079
Db	1026	tgatggtgcccagagaacgctggatctttgacggctcagtcctgggacctgtgagcaagac	1085	Db	1986	cctgtcatctacttccacaaagagaacggggttactgtcattacagataacagttc	2045
QY	421	agaaggttcccaagactccagaattgcccagacttcaatcctcagtcctcttccacc	480	QY	1080	atcctcagacattgtggagagctcaactcttatataaaatctcaaaactcttccattc	1139
Db	1086	agaaggttcccaagactccagaattgcccagacttcaatcctcagtcctcttccacc	1145	Db	2046	atcctcagacattgtggagagctcaactcttatataaaatctcaaaactcttccattc	2105
QY	481	ctctcagtggaatcgagaagaacagtagagtaactgggaatccaggggagtagagaatt	540	QY	1140	agagtattctctcttctcactcagactgagagaagtaaacatctcactatgaagg	1199
Db	1146	ctctcagtggaatcgagaagaacagtagagtaactgggaatccaggggagtagagaatt	1205	Db	2106	agagtattctctcttctcactcagactgagagaagtaaacatctcactatgaagg	2165
QY	541	cattgaaccatcacagaaataaattggacttaagcttcttcg-tggcaaatgattc	599	QY	1200	ggaatacgtcagcgccttctactgagtcgagtcctgtcatcacatccaaactctccgtcta	1259
Db	1206	cattgaaccatcacagaaataaattggacttaagcttcttcg-tggcaaatgattc	1265	Db	2166	ggaatacgtcagcgccttctactgagtcgagtcctgtcatcacatccaaactctccgtcta	2225
QY	600	cccaaccttggagaacatcagcttgcacagcagctctgaggtgcaaaatggaaagtcctat	659	QY	1260	cacacccacattaatatgcccgaacacttgcctctctctctctctctctctctctctt	1319
Db	1266	cccaaccttggagaacatcagcttgcacagcagctctgaggtgcaaaatggaaagtcctat	1325	Db	2226	cacacccacattaatatgcccgaacacttgcctctctctctctctctctctctctt	2285
QY	660	gtctcagactgagactgtcttaggtcagtcgcacccatagagagtgagagatcactgc	719	QY	1320	tagtgactt	1379
Db	1326	gtctcagactgagactgtcttaggtcagtcgcacccatagagagtgagagatcactgc	1385	Db	2286	tagtgactt	2345
QY	720	aaactggctttgaccaacagcaacatctgcagatgtgacaggaagctctcttccata	779	QY	1380	gacctgacctctgtgtcacaatcccaaccatttatttctcaattttacatcaaacag	1439
Db	1386	aaactggctttgaccaacagcaacatctgcagatgtgacaggaagctctcttccata	1445	Db	2346	gacctgacctctgtgtcacaatcccaaccatttatttctcaattttacatcaaacag	2405
QY	780	tcctgaagtgtaattgcttccagtggttaccagttctcagactctactgtacagctgg	839	QY	1440	ggcctctgtgcatctactaaagtctacactctgtatgcataccacacccatggcttctccacc	1499
Db	1446	tcctgaagtgtaattgcttccagtggttaccagttctcagactctactgtacagctgg	1505	Db	2406	ggcctctgtgcatctactaaagtctacactctgtatgcataccacacccatggcttctccacc	2465
QY	840	aggaagtcacacagcattggagataggagtatttaccagacttcttccatcttccctc	899	QY	1500	atacactttaccagatctcttaacagacatctacatctgccccacttctctctcaaac	1559
Db	1506	aggaagtcacacagcattggagataggagtatttaccagacttcttccatcttccctc	1565	Db	2466	atacactttaccagatctcttaacagacatctacatctgccccacttctctctcaaac	2525
QY	900	ggaagcttgaatttcacagcaccacgctggagaaacttcaa-----	940	QY	1560	aaacttggcacagctatcttctaccctctccctggccagggcgaaggagagactctgtgac	1619
Db	1566	ggaagcttgaatttcacagcaccacgctggagaaacttcaa-----	1625	Db	2526	aaacttggcacagctatcttctaccctctccctggccagggcgaaggagagactctgtgac	2585
				QY	1620	ttcatttcagacatcaacaatgacatcattcatgacaatgctccatagtagtcaaaactgc	1679
				Db	2586	ttcatttcagacatcaacaatgacatcattcatgacaatgctccatagtagtcaaaactgc	2645
				QY	1680	agaccttaagagccagagocccccacacccaagagaagtcattacagaaatcaaaagtcacc	1739
				Db	2646	agaccttaagagccagagocccccacacccaagagaagtcattacagaaatcaaaagtcacc	2705
				QY	1740	aagcctgggtgtctctgcccacagagtcccaacaaagctgttaacaaacactctcttggcc	1799

|||||
Db 2706 AAGCCTGGTGTCTCTGCCACAGTCCACCAAGCTGTAACAACAACACTCTCTCTTGCC 2765
QY 1800 tcaactcttaacagagtcctccacagagaaaccccttccagccacaagaacacactagc 1859
Db 2766 TCCATCTTTAACAGAGTCTCTCCACAGAGCAAAACCCCTCCAGCCACACAGCAACTTAGC 2825
QY 1860 acaaatgtctccaaacttcaaaactaccattctgaagacctctcagcccttatgaccac 1919
Db 2826 ACAAAATGTCTCCAACTTTTCAAACTACCAATCTCTGAAGACCTCTCAGCCCTTTATGACCAC 2885
QY 1920 tcttggcacctgtcaagcacagcatctctgttcaactggcccttatagccgtacagactac 1979
Db 2886 TCTTGGCACCTCTCAAGCACACATCTCTGTCTCACTGGCCCTATAGCCGTACAGACTAC 2945
QY 1980 agtggaaaacagctctcgtgacccatcctgtaaaactagtttccctcaaatctcaacaga 2039
Db 2946 AGCTGGAACACAGCTCTCGCTGACCCATCTCTGAAATCTAGTTTCCCTCAAACTCTCAACAGA 3005
QY 2040 agtggcatcagcacagaaggaacagagtgattgtggatgctaccactggattgatccc 2099
Db 3006 AGTGGCATCAGCACAGAAGAACCGAGTGATTTGGATGGCTACCACTGGATTTGATCCC 3065
QY 2100 tttagccagtgatccacatcagcaaaagaaatgacacaaagcttggggttacagcaga 2159
Db 3066 TTTGACCACTGTACCCACATCAGCAAAAGAAATGACACAAAGCTTGGCGTTTACAGCAGA 3125
QY 2160 gtacagccagcttcagttccctcgaaacatctctctcccaaacacacagttgttcc 2219
Db 3126 GTACAGCCAGCTTACAGTTTCCCTCGGAACATCTCTCTTCCCAACACACAGTTGTTTC 3185
QY 2220 cagggctgaagacttggctcccaaatctgcaaccttggctgttcagagcagcacagtc 2279
Db 3186 CACGGCTGAAGACTTGGCTCCCAAACTGCCACCTTTGCTGTTTCAGAGCAGCACAGATC 3245
QY 2280 accaaacaaactgtcctcttcagcctcagtcacacagctgtgtgtgaaccccttgttcca 2339
Db 3246 ACCAACAACACTGTCTCTTACGCTCAGTCAACAGCTGTGTGTGAACCCCTTGTCTTCA 3305
QY 2340 caatggcgaatcgttcgacagaaacaccagcgtgctaccactcaggtgcccgccttc 2399
Db 3306 CAATGGCGAATGCTCGCAGAACACACACCCGCTGCTACCACTGCAGGTGCCCCGCTTC 3365
QY 2400 ctggcaaggagtgattgcagtgatgtgaatgagtgctgtcgaacccctgcccac 2459
Db 3366 CTGGCAAGGGGATGATTGCAGTGTGATGTGAATGAGTGCTGTCTGCAACCCCTGCCATC 3425
QY 2460 cacagccagtgcaacaaactactcagggatccttatactgcaaatgcccgttgggtacca 2519
Db 3426 CACAGCCACGTGCAACAATACTCAGGGATCTTTATCTGCAAAATGCCCGTGGGTACCA 3485
QY 2520 gttggaaaaaggatgatgcaattgttagaaccttcgtgacagagtttaaatgaagag 2579
Db 3486 GTTGGAAAAAGGATATGCANTTTGGTTAGAACCTTCGTGCACAGAGTTTAAATTAAGAG 3545
QY 2580 aacttttcttaaatacaactgtgaaaaacattcagaccctacaagaagtgaataatgagat 2639
Db 3546 AACTTTTCTTAATAACAACGTGTGAAAAACATTCAGACCTTACAAGAGTTGAAAAATGAGAT 3605
QY 2640 caccaaaagtttaaatatgtgttttcagcgtttacactagttacatccgatctacagttca 2699
Db 3606 CACCAAAAGCTTAAATATGTGTGTTCAGCGTTTACCTAGTTTACATCCGATCTACAGTTCA 3665
QY 2700 cgcctctaggaggtccaaacgctgtgtatctcactgcaaaacacacctttccctggcctc 2759
Db 3666 CGCCTCTAGGGAGTCCACCGCGTGTGTATCTCAGTGCACAAACACCTTTTCCCTGGCCTC 3725
QY 2760 caatgtgacgtatttgactgtggtgataggaagcagaaatgtgtcaaacctctctgcaagtc 2819
Db 3726 CAATGTGACCTATTTGACCTGCTGATAGGATGCAGAAATGTCAAACTCTCTGCAAGTC 3785
QY 2820 ctctgtgaggtctgcagctcttggatctcagagggcgatctttagagcggcgacgtt 2879
|||||

Db 3786 CTCTGTCTGAGTCTGCCAGCTCTTGGGATCTCAGAGCGGATCTTTAGAGCGGGCAGCTT 3845
QY 2880 gtcaacagcgaagagtccegaatgtgacaaaagacacacctccatctgcaactgaacctggaacgg 2939
Db 3846 GTCAACAGCGGAAGAGTCCCGAATGTGACAAAGACACCTCTCATCTGCACTGACCTGGACGG 3905
QY 2940 cgttgcctctgcagtgcaagtgcgggatacttcaagtcaaaagaatggaacacactcctg 2999
Db 3906 CGTTGGCCTGTGCCAGTGCAGTGCAGGTCTGAGTTCAGTTCACCAAGATGACACACTCTCTG 3965
QY 3000 ccgagcatgtgaagatggatataggtctgaaaatgaaacctgcatgagttgcccatttgg 3059
Db 3966 CCGAGCATGTGAAGATGGATATAGGCTTGAATGAACCTGCAATGATGATGGCCATTTGG 4025
QY 3060 ccttgggtgctcaactgtgaaaacccctatacagcttatcactgtgtgtgctgcagccgc 3119
Db 4026 CTTTGGTGGTCTCAACTGTGAAACCCCTATCAGCTTTATCAGTGTGTGTGATGCGAGCCGC 4085
QY 3120 gggaggtgggctcctgctcactcctagtcacgcactgattgttacctgttgcagagaagaa 3179
Db 4086 GGGAGGTGGGCTCTCTCATCTAGGCATCGCACTGATTTGTTACCTGTTGCGAGAAAGAA 4145
QY 3180 taaaaatgacataagaacaaactccttccaaaagtggagatttccaaaatgtccccatctgc 3239
Db 4146 TAAAAATGACATGAACAAACTCATCTTCAAAAGTGGAGATTTCCAAATGTCCCCATATGC 4205
QY 3240 tgaataccctcaaaaactcctcgctcacaagaatggggccgagaagctattgaaatgcacga 3299
Db 4206 TGAATACCCCAAAATTCCTCGCTCACNAGANTGGGGCCGAGAGCTATTTGAAATGCATGA 4265
QY 3300 gaatggaagtaccaaaacactcctccagatgaaggatgtgtactactcgcctcacaagtgt 3359
Db 4266 GAATGGAAGTACCAAAAACCTCTCCAGATGACGGATGTGTACTACTCGCCTACAACTGT 4325
QY 3360 aaggaatccagaactgaaacgaacggactctaccoggcctacacactggaactgccagatc 3419
Db 4326 AAGGAATCCAGAACTTGAACGAACAGGACTCTACCCGGCTACACTGGACTGCCAGGATC 4385
QY 3420 acggactctctgatitcccgagacagataaccgctcttcatcagtgatgaagaacag 3479
Db 4386 ACGGCATCTTGTGATTTTCCCGACAGATATAACCGTCTTTTATCATGATGATGAAACAG 4445
QY 3480 aagaagagactactttaagtcaggagagagaggactcattgtctgtgagccagtcacc 3539
Db 4446 AAGAAGAGACTACTTTTAAGTCCAGGAGAGAGAGGACTCATTTGCTGTGAGCAGTCAAC 4505
QY 3540 tgggaactcgtctcaagagaccacacagagactgcccagagatttgcggagaccac 3599
Db 4506 TGGACCTCTGCTCAGAGAGGCCACACAGAGGCTGGCCCCAGGATTTGTGCGGAGGCCAC 4565
QY 3600 gctgagtggaagcagaagaggagacagggcatgcccgggctgacacagtggaaggagaca 3659
Db 4566 GCTGAGTGGCAACGACGAGAGGGACAGGATGCGGGGCTGACCAAGTGGAGGAGACA 4625
QY 3660 ggtggatgtggaaccacaggtcgtcattcagcaccttctgttactctgtgaacgtgaat 3719
Db 4626 GGTGGATGTGGAAACACAGGCTGCTATTTCAGCACCTTTTGTGTACTGTGAACGTGAAT 4685
QY 3720 gttgggacagtatcaagagactcctctgactgactgcacactgcacagccagcggcg 3779
Db 4686 GTGGGCGAGTATCAAGAGAGTCTCTCTGAGTGACTGACCATGGCATGGCATGGCAGGGGGC 4745
QY 3780 actattagccagggcagaccactagactcagtcagggagacctgtttccctctcgtttg 3839
Db 4746 ACTATTAGCCAGGACAGACCACTAGACTTCAGTGCAGGGACCTGGTTTCCCTTCGTTTG 4805
QY 3840 caacttagtaaaatgggtgggaggttcccttttggatctctgtttttgagactgttccagaaa 3899
Db 4806 CACTTTAGTAAATGGGTGGAGGTTTCCCTTTTGGATCTCTTTTGTAGACTGTTCACAGAAA 4865
QY 3900 gaaggcttctctcccgagagacacttccatagcagcaaatgtgtgatttcttctgcaacaa 3959
Db 4866 GAAGGCTTCTCTCCCGAGACACTTCCATAGGCAGCAATTTGGTGATTCATTTGCASCAA 4925
|||||

Qy	3960	aaactgctgtttaatttttctctgcccagcgctgctgctaaacaacagatgagga	4019
Db	4926	aaatactggctgtgttaatttttctctgcccagcgcctgctgctaaacaacagatgagga	4985
Qy	4020	tgaagctaccacactgaagctgaagatgtcgccattgaacgagcagtggttttcataatattt	4079
Db	4986	tgascgtaccacactgaagctgaagatgtcgccattgaacgagcagtggttttcataatattt	5045
Qy	4080	ctaggttgctctatgctacagtttccaaagccagccccacagtgtaggaaaatgtgtgaggc	4139
Db	5046	ctaggttgctctatgctacagtttccaaagccagccccacagtgtaggaaaatgtgtgaggc	5105
Qy	4140	accgcacacaactgcaatgtgttttttaagtcaaggtagacacatgatatttaagatattttt	4199
Db	5106	accgcacacaactgcaatgtgtgtttttaaagtcgaagtgacacatgatatttaagatattttt	5165
Qy	4200	tttaaaatctcttgcgattaaatctcaacttttcaaaagcctgtagcagggcaaaac	4259
Db	5166	ttttaaatactcttgcgattaaatctcaacttttcaaaagcctgtagcagggcaaaac	5225
Qy	4260	aaacttatattggcttttagctgagagctcagcagcgagatgcaggcaggggggcacttt	4319
Db	5226	aaacttatattggcttttagctgagagctcagcagcgagatgcaggcaggggggcacttt	5285
Qy	4320	tcatactgagggcccagcctggggcctgggactctgataccactctgagggccagagg	4379
Db	5286	tcatactgagggcccagcctggggcctgggactctgataccactctgagggccagagg	5345
Qy	4380	cagctcgatgagcagagaaatgtcaaaactgaacagcaggtttccaccactctaggaagca	4439
Db	5346	cmctcgctatgagcagagaaatgtcaaaactgaacagcaggtttccaccactctaggaagca	5405
Qy	4440	gcttgtagccctcgacgtggaatggtgttagagggatgggctgaaataggcaggttaga	4499
Db	5406	gcttgtagccctcgacgtggaatggtgttagagggatgggctgaaataggcaggttaga	5465
Qy	4500	ttctctgatacaacagtgctttgggaagctgtgtggaatcctgagggaagaacagggagcc	4559
Db	5466	tttctctgatacaacagtgctttgggaagctgtgtggaatcctgagggaagaacagggagcc	5525
Qy	4560	gaatgagccacacatgattgtctccacggctactcagcacttgatccacgaatct	4619
Db	5526	gagatgagccacacatgattgtctccacggctactcagcacttgatccacgaatct	5585
Qy	4620	catgtccaaaaacccatgtaaactttcaacactcaagctgtttatctggctgaagaa	4679
Db	5586	catgtccaaaaacccatgtaaactttcaacactcaagctgtttatctggctgaagaa	5645
Qy	4680	ataactttttttctcaccacagtcattgtacctcttcataatgctgtgtcgaccctcc	4739
Db	5646	ataactttttttttctcaccacagtcattgtacctcttcataatgctgtgtcgaccctcc	5705
Qy	4740	agaaacgtgttatacttcacgtcagtgtaggaactgaagacttcggttggtcgagg	4799
Db	5706	agaaacgtgttatacttcacgtcagtgtaggaactgaagacttcggttggtcgagg	5765
Qy	4800	aaactgaggttgaccttcgggaaggaagtccactactcttattattatgtcctgtgatg	4859
Db	5766	aaactgaggttgaccttcgggaaggaagtccactactcttattattatgtcctgtgatg	5825
Qy	4860	tgggtcctgcaggagacatccagtaactcgtgtgtctttaaattgccacctggggaactgt	4919
Db	5826	tgggtcctgcaggagacatccagtaactcgtgtgtctttaaattgccacctggggaactgt	5885
Qy	4920	gtttattggcctctcttgggggcatcctgggttttgatgaagtgggggaatacaagagta	4979
Db	5886	gtttattggcctctcttgggggcatcctgggttttgatgaagtgggggaatacaagagta	5945
Qy	4980	aaagaattgtctcacctgaacggggagtcctcgcttccatctcggaaatggtgcag	5039
Db	5946	aaagaattgtctccacctgaacggggagtcctcgcttccatctcggaaatggtgcag	6005

Qy	5040	ccactggggacagtcttgcgccggggcatggttggttcttcaagtgctcctcaataataatc	5099
Db	6006	CCACTGGGGACAGTTCTGCCCCGGGCATGGTTGTTCTTCAAGGTCCCTAAATATAATC	6065
Qy	5100	cctattcttacaataactcttggccctgatgggttttaagcaagaactcctgtgtcccatgg	5159
Db	6066	CCTATTCTTACATAAATCCTTGGCCCTCGATGGTTTAAAGCAAGAACTCCTGTGTCCMATGG	6125
Qy	5160	tctccaccactcaccatccactcgtgtgtagcaagagtcctagtcaggggaggtgcatttt	5219
Db	6126	TCTCCACCACCTCACCATCACCCCTGCTGTAGCAAGAGTCTCTAGTCAGGGGAGGTGCATTTT	6185
Qy	5220	agtagttaattcactattatccatgagagataataaaaaagagaaactgtttttatcatcagtgga	5279
Db	6186	AGTAGTTACATTCGCATTAATCCATGAGATAAATAAAAGGAGAVCTGTTTTTATCAGTGGGA	6245
Qy	5280	ggctaaactaaaatttcaaaagtgcgccttttggaaactcttgggcctctctctctgtaga	5339
Db	6246	GGCTAACCTTAAATTTCAAAGTGTCCCTTTTGGAAATCTTGGGCCTCTCTCTCTGTAGA	6305
Qy	5340	accaatggcccttgtggctcagcgccctgcacctaactcgagagttcttgagctcctgcga	5399
Db	6306	ACCAATGCCCTTTGTGGCTCAGCGGCTCGCACCTTAACCTGAGAGACTTCTGAGCTCCTGCA	6365
Qy	5400	gtcacctgagccacacagactaggtcttcttggctcctctccgc	5441
Db	6366	GCTCACCTGAGCCACAGACTAGGCTTCTTGGCTCCTTCCCG	6407
RESULT 5			
US-08-944-496-7			
; Sequence 7, Application US/08944496			
; Patent No. 612433			
; GENERAL INFORMATION:			
; APPLICANT: FALB, DEAN A			
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE			
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE			
; NUMBER OF SEQUENCES: 54			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: PENNIE & EDMONDS LLP			
; STREET: 1155 Avenue of the Americas			
; CITY: New York			
; STATE: New York			
; COUNTRY: USA			
; ZIP: 10036-2711			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/944,496			
; FILING DATE: 06-OCT-1997			
; CLASSIFICATION: 514			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 08/599,654			
; FILING DATE: 09-FEB-1996			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 08/485,573			
; FILING DATE: 07-JUN-1995			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 08/386,844			
; FILING DATE: 10-FEB-1995			
; ATTORNEY/AGENT INFORMATION:			
; NAME: CORUZZI, LAURA A			
; REGISTRATION NUMBER: 30,742			
; REFERENCE/DOCKET NUMBER: 7853-104			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (212) 790-9090			
; TELEFAX: (212) 869-8864			
; TELEX: 66141 PENNIE			
; INFORMATION FOR SEQ ID NO: 7:			
; SEQUENCE CHARACTERISTICS:			

;
; LENGTH: 6407 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cdna
; US-08-944-496-7

Query Match 73.3%; Score 5098.4; DB 3; Length 6407;
Best Local Similarity 94.3%; Pred. No. 0;
Matches 5412; Conservative 21; Mismatches 8; Indels 301; Gaps 2;

Qy 1 gttcgtgaagaattgccctcttcttcaaaacaaagatggaacagcctcgagatgggaac 60
Db 666 GTTCGATGAAGAATTCGCCGCTTTTCAACAAACAGAGTGAACAGCCTCGGAGATGGGAAC 725
Qy 61 agagagggcagatgggggtgtcagaagaatggactgtcacagccaaagagccaccacttc 120
Db 726 AGAGAGGGCGATGGGGGTGTGAGAAGAAATGGACTGTGCACAGCCAAAGAGCCACCACTTC 785
Qy 121 ggettgagccctcccttctctctgttggagatggagagctgaccgccttctag 180
Db 786 GGCCTGGAGCCCGTCTTCTTCTGCTTGGAGATGGGAGAGTGACCAAGCCCTTCTAG 845
Qy 181 gaagagaaattcctcaggaacagatctctctggtcatttctacaggaagcagcttc 240
Db 846 GAAGAGAAATTCCTCAGGACAGATCTCTCTGGCTGCATTTCTACAGGACAGCAGCTTC 905
Qy 241 ctctcctctctagaccttctccacacctctgaaagtacagagaagcttaacaactccac 300
Db 906 CTCCTCTCTTAGACCTTTCCTCACCTCTTGAAAGTACAGAGAAGCTTAACAACCTCCAC 965
Qy 301 tggcctccagagctcctcagtcagtcacaaacaaagacaatgcattgttctaccggttcac 360
Db 966 TGGCCTCCAGAGCTCCTCAGTCAGTCAACAAACAGACAATGCATGTTGCTACCGTGTTCAC 1025
Qy 361 tga tgg tgg ccc gagaac agctg agatctt tga cgg tgc agtct gga cct gtg ag ca ag ac 420
Db 1026 TGATGTTGGCCGAGAACGCTGGCATCTTTGACGGTCAGTCTGGGACCTGTGAGCAAGAC 1085
Qy 421 agaaggcttcccaagactccagaattgccagactctccagactctcagctcttcttcaac 480
Db 1086 AGAAGGCTTCCCAAGGACTCCAGAATTGGCAGACTTCATCTCTCAGTCTCTTTTCCACC 1145
Qy 481 ctctgcagtggaatcgaagaagaacagtagtaactgggaatccagggtgaggaatt 540
Db 1146 CTCCTGCAGTGGAAATCGAAGAAACACTAGATTAACCTGGGAATCCAGGGGATGAGGAATT 1205
Qy 541 cattgaaccatccacagaagaatgaatttgacttaactcctttgcg - tgg caaaatgattc 599
Db 1206 CATTTGAACCATCCACAGAAAATGAATTTGGACTTACGCTTTGCGTTTGGCAAAATGATTC 1265
Qy 600 cccaacctttggagaacatcagcttgcagcagctctgaggtgcaaatggaagtcccat 659
Db 1266 CCCAACCTTTGGAGAAGCATCAGCTTGGCAGCAGCTCTGAGGTGCAAAATGGAAGTCCCAT 1325
Qy 660 gtctcagactgagactgtctaggttcagtcgacccatgagagtggaagatcactgc 719
Db 1326 GTCTCAGACTGAGACTGTGCTAGGTGAGTGCACCCCATGAGAGGTGGAGAGATCACTGC 1385
Qy 720 acactggcttctgacaaacgacaacatctgcagatgagaggaagctctgtctcata 779
Db 1386 ACATGGCTCTTGACCAACAGCAACATCTGCAGATGTGACAGGAAGCTCTGCTTCATA 1445
Qy 780 tcttgaagggtgaattcagttcagttgtgacccagttctcagactctacttacagctgg 839
Db 1446 TCCTGAAGGTGTGAATGCTTCAGTGTGACCCAGTTCCTAGACTCTACTGTACAGTCTGG 1505
Qy 840 aggaagtcaacagcatgtggagataggagtatttcagagctcttctacatcttctctc 899
Db 1506 AGGAAGTCAACAGCATTTGGAGATAGGAGTTATTTCAGAGCTCTTCATCTACATCTTCCTC 1565
Qy 900 ggaagcttgaaattcatcagcacaacgaogtggagaacgttcaa----- 940

Db 1566 GGAAAGCTTGAATTCATCAGCACACCTGGAGAACGTTCAACCTTGAAGACACCGGAGA 1625
Qy 941 ----- 940
Db 1626 GCCAGGCCAAGCACTAGGTGACAGTTCGCCAATGSCAGAGACAGGACTTCTTGGGGTGCC 1685
Qy 941 ----- 940
Db 1686 CTCTCTCGGCACCCACACACTTGGCTACTGTCTACTTGGAAACGGGGAACGCACACTGCGGTC 1745
Qy 941 ----- 940
Db 1746 TGTCAACCTTCACCAACACCAGCATGACGAGCTTCTGGGGAAGCAGGAGCCCTGCAGC 1805
Qy 941 ----- 940
Db 1806 GGCCATGCCCCAAAGAAACAGAGGGTGCTCTCTGCACGTAAACGTGACGAGACATGGG 1865
Qy 941 -----tcgctgggattagctacgg 959
Db 1866 CTTGGTCTCACGGTCACTGGCCGCTCCAGTGCACCTCGGAGTCTGCTGGGATTAGCTACGG 1925
Qy 960 tcaagtcgtggaacagctattgaacaaagagacttccagagaccacacagaccacacta 1019
Db 1926 TCAAGTCGTGGCAGACGCTATTGAACAAAGGACTTCCAGGACACACAGACCACACCTA 1985
Qy 1020 cctgtctctacttctccaaagagacggcggttactctccattacagataacagttc 1079
Db 1986 CTTGTCTACTTTCACAAAGGAGAACGGCGCTTACTGTCCATTACAGATAACAGTTC 2045
Qy 1080 atctcagacattgtggagagctcaactcttattataaatactcaaacctcttcacattc 1139
Db 2046 ATCCTCAGACATTTGTGAGAGCTCAACTCTTATATTAAAAATCTCAAACTCTTCACATTC 2105
Qy 1140 agagtatctctctctctctcagtcgaactgagagaagtaaacatctcatctctatgacgg 1199
Db 2106 AGAGTATTCCTCTCTTCTCATGCTCAGACTGAGAGAAGTAACATCTCATCTATGACGG 2165
Qy 1200 ggaatagctcagcctctctcagtcgagtcagctctcagacacacacacacacacacac 1259
Db 2166 GGAATATGCTCAGCCTTCTACTGAGTCGCCAGTTCTGCATACATCCAACTTCCTGCTCCTA 2225
Qy 1260 cacaccaccatttaataatgcgaacacttcggtgtgtctctggacactgatgctgagttgt 1319
Db 2226 CACACCCACCATTAATATGCCGAACACTTCGGTTGTTCTGGACACTGATGCTGAGTTGT 2285
Qy 1320 tagtgaact 1379
Db 2286 TAGTGACT 2345
Qy 1380 gctctgcccctctgtgtcacaatcccaccatttattttcatcaatttaccatcaaccag 1439
Db 2346 GCCTCTGCCCTCTGTGTGCACAATCCCACTCATTAATTTTCATCAATTTTACATCAACACAG 2405
Qy 1440 ggcctctgtgcatctactataaagtctacctgtgatgcacacacacacacacacacacac 1499
Db 2406 GGCTCTGTGCATCTACTATAAGTCTACCTCTGTATGCATCCACACATGCTTCTCTCCTCACC 2465
Qy 1500 atcaactttaccagatcctttaaogacatctacatctgccccacacttctctctcaacaaac 1559
Db 2466 ATCACCCTTTACCAGTATCTTAAACGACATCTACATCTGTGCCCCACTTCTGTCTCACAAAC 2525
Qy 1560 aacctgcccacagctcattcttaccctctgcttcccagggcagggagagactcctgtgac 1619
Db 2526 AACCTTGCCACAGTCATCTTCTACCCCTGTCTGCCCGAGGAGGAGACTCTCTGTGAC 2585
Qy 1620 ttcatctcagacatcaacaatgacatctcatgacaatgtctccatagtagtcaaacatgc 1679
Db 2586 TTCAATTTGACAGATCAACAATGACATCATTCATGACAATGCTCCATAGTAGTCAAACTGC 2645
Qy 1680 agaccttaagagccagagccccacacacacacacacacacacacacacacacacacacac 1739

Db 2646 AGACCTTAAGAGCGCAGAGCACCCACACACCAAGAGAAAGTCATTACAGAATCAAAAGTCAACC 2705
QY 1740 aagcctggtgtctctgccacagagtcaccacaaagctgtacacaaactctctcttgcc 1799
Db 2706 AAGCCTGGTGTCTCTGCCACAGAGTCACCAAAAGCTGTAAACAAACACTCTCCTTTGGCC 2765
QY 1800 tccatctctaacagagtcctccacagagcaaacctctccagccacacagcaacaaacttagc 1859
Db 2766 TCCATCTTAACAGAGTCCTCCACAGAGCAAAACCTTCCAGGCCACAGCAACCAACTTAGC 2825
QY 1860 acaaatgtctccaactttcacaaactaccattcttgaagacacctctcagcctcttatgaccac 1919
Db 2826 ACAAAATGCTCCAACTTTCAACAATACCACTTCTGAAGACCTCTCAGCCTCTTATGACCAC 2885
QY 1920 tccctggcacctgttcaagacagcatctctggttcacttgccctatagccgtacagactac 1979
Db 2886 TCCCTGGCACCCCTGTCAAGCACAGCATCTCTGGTCACTGGGCCCTATAGCCGCTACAGACTAC 2945
QY 1980 agctggaaaaacagctctcgtgacccatcctgaaataactagtctcctcaaatctcaacaga 2039
Db 2946 AGCTGGAAAACAGCTCTCGCTGACCCATCTCTGAANAATAGTTCTCCTCAAAATCTCAACAGA 3005
QY 2040 agtggcatcagcacagaaaggaaacccagtgattgttgatgcttaccactgattgatccc 2099
Db 3006 AGGTGGCATCAGCACAGAAAGGAACCCAGTGTATTGGATGCTTACCACCTGGATTGATCCC 3065
QY 2100 ttgtaccagtgtaaccacatacagcaaaagaataagcacaaaagcttgggttacagacaga 2159
Db 3066 TTTGACCAGTGTACCCACATCAGCAAAAGAAATGACCACAAAGCTTGGCGCTTACAGACAGA 3125
QY 2160 gtacagcccagcttcagttccctcctcggaacatcctctctcccaaacacacagttgttc 2219
Db 3126 GTACAGCCGAGTTTCAGTTTCCCTCGGAACATCTCCCTTCTCCCCAAACACAGTTGTTTC 3185
QY 2220 cacggtggaagacttggctcccaaatctgcacacttctgcttgcagcagcagcacagtc 2279
Db 3186 CAGGCTGAAGACTTGGCTCCCAAAATCGCCACCTTTGCTGTTTCAGAGCAGCACACAGTC 3245
QY 2280 accaacaacactgtctctcagcctcagtcacacagctgtgctgtgaacacctgtcttca 2339
Db 3246 ACCAANAACACTGTCTTCAGCCTCAGTCAACAGCTGTGCTGTGAACCCCTGTGCTTCA 3305
QY 2340 caatggcgaatgctgcagacacacacccgctgggtaccactgcaagtgccgccttc 2399
Db 3306 CAATGGCGAATGCTGCGCAGACAAACACCGCTGGCTACCAGTGCAGGTGCCCGCCTTC 3365
QY 2400 ctggcaagggaatgtatgcagtggtggtatgtgaatgagtgctgtcgaaacccctgcccac 2459
Db 3366 CTGGCAAGGGGATGATTGCAAGTGTGGATGTGAATGAGTGTGCTGTGCAAAACCCCTGCCCATC 3425
QY 2460 cacagccactgcaacaataactcaggtatcctttatctgcaaatgcccggttgggtacca 2519
Db 3426 CACAGCCACGTGCAACAAATACTCAGGGATCCTTTATCTGCAAAATGCCCGGTGGGTACCA 3485
QY 2520 gttggaaaaaggatatgcaatttggtagaacctctgtgacagagtttaaatataagag 2579
Db 3486 GTTGGAAAAAGGATATGCAATTTGGTTAGAACCTTCGTGCACAGAGTTTAAATTAAGAG 3545
QY 2580 aactttcttaatacactgtgaaaaacattcagacctcaagaagtgtgaataatgagat 2639
Db 3546 AACTTTTCTTAATACAACTGTGAAAAACAAATTACAGACTTCAAGAAGTTGAAAAATGAGAT 3605
QY 2640 caccacacgttaaatatgtttttcagcgttaccttagttacatcgatctcagttca 2699
Db 3606 CACCAAAACGTTAAATATGTGTCTTTCAGCGTTACCTTAGTTTACATCCGATCTACAGTTCA 3665
QY 2700 cgcctctaggaggtccaaacgggtgggtgatctcactgcaacaacacctttccctggcctc 2759
Db 3666 CGCCTTAGGAGTCCAAACGGGTGGGTGATCTCACTGCAACAACACCTTTTCCCTGGCCTC 3725
QY 2760 caatgtacgctatttgacacctggtgataggaatgcagaaatgtgtcaacctcctgcaagtc 2819
Db 3726 CAATGTGACGCTATTGTGACCTGTGATAGGATGCAGAAATGTGTCAACTCCTTGCAAAGTC 3785

QY 2820 ctctgctgaggtgtccagctcttgggatctcagagggagatcttttagagcggcagctt 2879
Db 3786 CTCTGCTGAGGTCTGCCAGCTCTTTGGGATCTCAGAGCGGATCTTTAGAGCGGCAGCGTT 3845
QY 2880 gtgcaacgggaagagtcctccgaatgtgacaaagacacctcatctgcaactgacacctggacgg 2939
Db 3846 GTGCAACGGGAGAGTCCCGAATGTGACAAAGACACCTCCATCTGCACGTGACCTGGACGG 3905
QY 2940 cgttgccctgtgcagtgcaagtcgggatactttcagttcaacaagatgggaccactcgt 2999
Db 3906 CGTTGCCCTGTGCCAGTGCAGTCCGGATACTTTTCAGTTCAACAAGATGGACCACTCCTG 3965
QY 3000 ccgagcagtgtgaagatggatataggttgaaaaatgaaacctgcatgtgagttgcccatttgg 3059
Db 3966 CCGAGCATGTGAAGATGGATATAGGCTTGAANAATGAAACCTTGCATGTGTCCTCATTTGG 4025
QY 3060 ccttgggtgtctcaactgtgaaaaacccctatcagcttctcactgtgtgtgacgcagccgc 3119
Db 4026 CCTTGGTGGTCTCAACTGTGGAAACCCCTATCAGCTTATCACTGTGTGTGATCGCAGCGC 4085
QY 3120 gggaggtgggtcctcgtcctcatccttaggcacgcactgattgttacctgtgtgcagaaagaa 3179
Db 4086 GGGAGGTGGGCTCCTGCTCATCTCCTTAGGCATCGCACCTGATTGTTACCTGTTGCAGAAAGAA 4145
QY 3180 taaaaatgaacataagaacaaactcatcttcaaaagtggagatttccaaatgtccccatatgc 3239
Db 4146 TAAAAATGACATAAGCAAACTCATCTTCAAAGTGGAGATTTCCAAATGTGCCCATATGC 4205
QY 3240 tgaatacccccacaaactcctcgtcacagaatggggccgagaaactattgaaaatgcataga 3299
Db 4206 TGAATACCCCAAAAANTCCTCGCTCAAGAATGGGGCCGAGAACTATTGNAATGTCATGA 4265
QY 3300 gaatgggaagtaccacaaacacctctccagatgaecggaatgtgtactactcgcctacaagtgt 3359
Db 4266 GAATGGAAGTACCAAAAAACCTCTCCAGATGACGGATGTGTACTACTCGCCTACAAAGTGT 4325
QY 3360 aaggaaatccagaaacttgaacgaaaggactctacccggcctacactggactgcaggatc 3419
Db 4326 AAGGAATCCAGAACTTGAACGAAACGGACTCTACCCGGCCTTACACTGGACTTGCAGGATC 4385
QY 3420 acggcatcttctgatttccccggacagtaaaacccgtcttctcatcagtgatgaaagcag 3479
Db 4386 ACGCATCTTTCGATTTTCCCGGACAGTATAACCCGCTCTTTCATCAGTGTGAAGCAG 4445
QY 3480 aagaagagactacttttaagtccaggagagagggactcaattgtctctgagccagtcacc 3539
Db 4446 AAGAAGAGACTACTTTTAAAGTCCAGAGAGAGAGGGACTCATTTGCTCTGTAGCCAGTCAAC 4505
QY 3540 tgggacctctgctcagagagaccgacccagaggtgcgcccaggatttgcggagagcac 3599
Db 4506 TGGACCTCTGCTCAGAGGACCCGACCCAGGAGGTGCGCCAGGATTTGTGGGAGGCCAC 4565
QY 3600 gctgagtggaacagagagggagcagcagcagcagcagcagcagcagcagcagcagcagcagc 3659
Db 4566 GCTGAGTGGCAACAGCAGAGAGGGACAGCATGCGGGCGTGCACCACTGAGGAGGAGACA 4625
QY 3660 ggtggaatgtggaacacacagcgtcctcattcagcacccttgttctgttactgtgaacgtgaat 3719
Db 4626 GGTGGATGTGGAACACACAGGCTGCTCATTTACGACACCTTTTGTGTTTACTGTGAACGTGAAT 4685
QY 3720 gtagggccagtatcaagagagtcctctgagtgactgcacatgagcactgcacccagggcg 3779
Db 4686 GTGGCCAGTATCAAGAGAGTCTCTGTGAGTACTGACCATGACATGACATGACACCGGGCG 4745
QY 3780 actattagcaggggcagacccactagacttcagtcaggggacctgtgttccctctcgcttg 3839
Db 4746 ACTATTAGCCAGGACACACACTAGACTTCAGTGCAGGACCTGGTTTCCCTTCGTTG 4805
QY 3840 cactttagtaaatgtgggtggaggttcccttggatctgtttagactgtttagactgttccagaaa 3899
Db 4806 CACTTTAGTAATTTGGGTGGAGGTTTCCCTTTTGGATCTGTGTTTGTGAGACTGTTCAGAAA 4865

Qy 3900 gaaggcttccttcccagagacacttccatagggcagcaatttgggtgattcatttgcagcaa 3959
Db 4866 GAAGGCTTCCTTTCCCGAGACACTTCCATAGGCAGCAATTTGGTGATTCATTCASCAA 4925
Qy 3960 aactagctgtgtaattatttcttcccagcgcctcgctgctaacaacagatgagga 4019
Db 4926 ANTACTGCTTGTAATATTTTCTGCCACAGCRCTGCGTGTCTAAACAACAGATGAGGA 4985
Qy 4020 tgagctgaccactgaagctgaagatgtcgccattgaacgacagtggttttcataatgttt 4079
Db 4986 TGASCGTACCACCTGAAGTCTGAAGATGTCGCCATTGAACGGACAGTGTTCATATGTTT 5045
Qy 4080 ctaggttgtcttatgtcacagtttccaaagcccccacagtgaggaatgtgtgagcc 4139
Db 5046 CTAGGTTGTCTTATGCTACAGTTCATGAGGTCAGCCACAGCTGAGGAAATGTGTGAGC 5105
Qy 4140 accgcacacactgcaatgtgtttttaaagtcgaagtgacacatgtatttaagattttt 4199
Db 5106 ACCGCACACAACTGCAATGTGTGTTTAAGTCAAGGTGACACATGTATTTAAGATTTTTT 5165
Qy 4200 tttaaaatctcttctgcagtttaaaatctcaccttttccaaacagccctggatcagggcaaac 4259
Db 5166 TTTAAATCTCTTTCAGTTCATCTCACCTTTTTCAAACAGGCTGATCAGGGCAAAAC 5225
Qy 4260 aactatatttggtttttagctgaggtctcagcagcgagatgcaggggggacattt 4319
Db 5226 AACTTATATVTGTTTGTAGCTGGAGGCTCAGCAGGCAGATTGCAAGGAGGGGGGCACTTT 5285
Qy 4320 tcacctaggggcccagcctgggctggagctcgtacacatgtgagggccagagg 4379
Db 5286 TCATCCATGAGGGCCCCAGCTGGGGCTTGGACACTCTGATCACCATTGTGAGGGCCAGAGG 5345
Qy 4380 cagctgcgtatggaggagaaatgtcaaaactgaacgcaggtttccaccactctaggaagca 4439
Db 5346 CAMCTGCGTATGGAGGAAATGTCAAACTGAACGCAGGTTTCACCACTCTAGGAAGCA 5405
Qy 4440 gcttgtgagccctcgcagctggaatgctggttagagggatgggctgaatagcaggttaga 4499
Db 5406 GCTTGTGACCCCCCTGCASCTGGATGTGTGTAGAGGATGGCTGGAATAGSCAGGTTAGA 5465
Qy 4500 ttctcgcatacaacagtcgttgggaagctgtgtggattcctgaggaagacagggagcc 4559
Db 5466 TTTCTCTGCATCAACAGTGTCTTGGGAASCTGTGTGGATTCTCTGAGGAAGAACAGGGAGCC 5525
Qy 4560 gagatggagccacacatgagtttgcctcacgcggtactgcagcactttgttaccacgaatct 4619
Db 5526 GAGATGAGCCACACATGAATTCCTCACCCTGCTACTGCAGCACTTTGTACCCAGATCT 5585
Qy 4620 catgtccaaaaccccatgtaaactttcaaccactcaaaactgatttattcggtgaagaa 4679
Db 5586 CATGTCCACAAACCCCATGTAACTTTCAACCACTCAAAAGSTGTTTATTCGGCTGAAGAA 5645
Qy 4680 ataacttttttctccaccagcattgtaccttttcataatggctgtgcgacccctcc 4739
Db 5646 ATAACTTTTATTTCTCACCAGTCATTTGTACCTCTTCATATGGATGTGCGACCCCTCC 5705
Qy 4740 agaaactggttatacttccagtcagtggtgggagaactgaagcttcggtgtggtcaggg 4799
Db 5706 AGAAACGTGTTATACTKCCAGTCAGTGTGGGAGAACTGAAGACTTCGGTTCAGG 5765
Qy 4800 aactgaggggtgaccttcgggaaggaagttccactcatcttattattatgcctgtgatg 4859
Db 5766 AACTGAGGGTTGACCTTCGGGAAGGAAGTTCCACTCATCTTATTTATTATGCTGTGATG 5825
Qy 4860 tgggtctgcagggagacatccagttactcgtgtctttaaattgacacccctggggaactgt 4919
Db 5826 TGGTCTCTCCAGGGAGACATCCAGTACTCGGTGTCTTKTAATTGCCACCTGGGGAACTGT 5885
Qy 4920 gtttattggcctctcttggggcatcctcgttttggatgaagtgagggaatacacaggtta 4979
Db 5886 GTTTATTGGCTTCTTTGGGGCATCTTGKTTCCGATGAAGTGAAGGGAATACAGAGGTA 5945
Qy 4980 aaagaattgtctccacccctgaagcgggagtcctcgtttcacatttctcggaaatggtgcag 5039

Db 5946 AAAGAATTTGTCACCCCTGAAGCGGGAGTCCCGCTTCACATTTCTGGAATGGTGCA 6005
Qy 5040 ccaactggggacagttctgccccgggcatggtgtgtttcttcaaggtcctcctaaataatc 5099
Db 6006 CCACCTGGGACAGTTCTGCCCGGGCATGTTGTTTCTTCAAGGTCCTCTAAATAATC 6065
Qy 5100 cctattctacataatccttggccctgatgggtttaagcaagaactcctgtgtcccatgg 5159
Db 6066 CTTATTCTTACATTAATCTTGGCCCTGATGTTTAAAGCAAGAACTCCTGTGTCMATGG 6125
Qy 5160 tctccacactcaccaatccacctcgtgtagcaagtgctctcagtgaggggaggtgcattt 5219
Db 6126 TCTCCACCACTCACCATCCCTGCTGTAGCAAGAGTCTTAGTCAGGGAGGTGCATTTT 6185
Qy 5220 agtagttaaattgcacttatccatgagataaataaaaggagaaactgttttatcagtgga 5279
Db 6186 AGTAGTTACATTTGCACATTATCCATGAGATAAATAAAGGAGAVCTGTGTTTATCAGTGA 6245
Qy 5280 ggcctaaactaaaatttcaaaagtgcgcctttttgaaactcttgggacctctctctctaga 5339
Db 6246 GGCTAACTAAATTTCAAAGTGTGCTTTTGAATCTTTGGGCTCTCTCTGTAGA 6305
Qy 5340 accaatggcccttctgtggtcacgcgctcgcacctaactgaagaggttctgagctcctgca 5399
Db 6306 ACCAATGCCCTTTGTGTGCTCACGGCTCGCACCTTAACCTGAGAGTTCCTGAGCTCCTGCA 6365
Qy 5400 gctcacctgagccacacagactagcttcttgcctccttcctccgc 5441
Db 6366 GCTCACCTGAGCCACAGACTAGGCTTCTTTGGCTCCTTTCCGC 6407

RESULT 6
US-09-328-111-83/c
; Sequence 83, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 83
; LENGTH: 584
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(584)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-83

Query Match 6.8%; Score 470.4; DB 4; Length 584;
Best Local Similarity 95.9%; Pred. No. 4.6e-121;
Matches 544; Conservative 0; Mismatches 16; Indels 7; Gaps 6;

QY	272	gaaagtagagagaagcttaacaactcca--ctggctctccagagctctctgaagtcagtcacaa	329
Db	567	GCAGGTNCAGAGAAGCTTAACAANTCCNACTGGGCTCCAAAGCTCCTNAGTCAGTCAA-	507
QY	330	caaagacaatgcattgtctacogtctcaactgatgttgcccgagaacgtgc-gatctc	388
Db	508	CAAAGCCAAATGCATGTGTGCCCCGCTGTTCACTGAAGTGGCCCGAAGACGCTGCGGATCT	449
QY	389	ttgac-ggtcagctctggagacctgtgagcaagacagaaggcttccccaaaggactccagaat	447
Db	448	TTGACGGGTCAAGTCTGGAACCTGTGAGCAAGACAGAAGGCTTCCCCAAGGACTCCAGAAT	389
QY	448	tgcac-gacttcactcactcagtcctctcttccactctcagctggaatcgagaagaaca	506
Db	388	TGTACGGACTTCATCCTCAGTCTCTTCTTTCACCTCTGACGTGGAAATCGAAGAANCA	329
QY	507	gtagagtaactgggaatccagggaatgaggaattcatigaacatccacagaaaaatgaat	566
Db	328	GTAGAGTAACGTGGGAATCCAGGGGATGAGGAATTCATTGAACCATCCACAGAAAAATGAAT	269
QY	567	ttagacttacgtctttgcg-tggcaaaaatgattccccaaacctttggagaacatcagcttg	625
Db	268	TTGGACTTAGTCTTTTTCGTTGGCAAAATGATTCCCCAACCTTTGGAGACATCAGCTTG	209
QY	626	ccagcagctctgaggtggcaaaatggaagtcctcattgtctcagactgagactgtgtctaggt	685
Db	208	CCAGCAGCTCTGAGGTGCAAAATGGAAGTCCCATGTCTCAGACTGAGAGTGTGTCTAGGT	149
QY	686	cahtgcaccccatgagaggtggagagatcaactgcacactggctcttggaccaacagacaa	745
Db	148	CAGTGCACCCATGAGAGGTGGAGAGATCACTGCACACTGGCTCTGTGACCAACAGACAA	89
QY	746	catctgcagatgtagcagaagctcgtctcatctcctgaaggtgtaatgcttcagtg	805
Db	88	CATCTGCAGATGTGACAGGAACTCTGCTTTCATATCCTGAAGGTGTGAATGCTTCAGTGT	29
QY	806	tgaccagttctcagactctactgtac	832
Db	28	TGACCCAGTTCTCAGACTCTACTGTAC	2

RESULT 7
US-08-480-994-7
; Sequence 7, Application US/08480994
; Patent No. 5834248
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,994
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:

```

; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-033
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 128 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
US-08-480-994-7

Query Match 1.8%; Score 126.4; DB 2; Length 128;
Best Local Similarity 99.2%; Pred. No. 1.2e-25;
Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps

QY 3120 gggaggtgggcctgctcatcctagcactgcactgagctgttaccctgttcagagaagaa 3179
Db 1 GGGAGGTGGGCCTGCTCATCTCCTAGGCATCGACTGATGTACTCTGTTCAGAAAGAA 60

QY 3180 laaaaatgacataaagcaaaactcatcttcaaaagtgagagattccccaatgtcccccatgc 3239
Db 61 TAAAAATGACATAAGCAAACTCATCTTCAAAAGTGAGAGATTTCCAAATGTCCCGGTATGC 120

QY 3240 tgaatacc 3247
Db 121 TGAATACC 128

RESULT 8
US-08-485-573-7
; Sequence 7, Application US/08485573
; Patent No. 5968770
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,573
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 128 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
US-08-485-573-7

Query Match 1.8%; Score 126.4; DB 2; Length 128;
Best Local Similarity 99.2%; Pred. No. 1.2e-25;
Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3120 gggaggtgggctctctatccttaggcacgcacgtgattgttacctgttgccagaaagaa 3179
Db 1 GGGAGGTGGGCTCTCTCTATCTCTAGGCATCGCACGTGATTGTTACCTGTTGCAGAAAGAA 60

QY 3180 taaaaatgacataagaacaaactctcttcaaaaagtggagatttccaaatgtccccatattgc 3239
Db 61 TAAAAATGACATAAGCAAACTCATCTTCAAAAGTGGAGATTTCCAAATGTCCCGGTATGC 120

QY 3240 tgaatacc 3247
Db 121 TGAATACC 128

RESULT 9
US-08-925-743-7
; Sequence 7, Application US/08925743
; Patent No. 6054558
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/925,743
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/485,573
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 128 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
US-08-925-743-7

Query Match 1.8%; Score 126.4; DB 3; Length 128;
Best Local Similarity 99.2%; Pred. No. 1.2e-25;
Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3120 gggaggtgggctctctatccttaggcacgcacgtgattgttacctgttgccagaaagaa 3179
Db 1 GGGAGGTGGGCTCTCTCTATCTCTAGGCATCGCACGTGATTGTTACCTGTTGCAGAAAGAA 60

QY 3180 taaaaatgacataagaacaaactcttcaaaaagtggagatttccaaatgtccccatattgc 3239
Db 61 TAAAAATGACATAAGCAAACTCATCTTCAAAAGTGGAGATTTCCAAATGTCCCGGTATGC 120

QY 3240 tgaatacc 3247
Db 121 TGAATACC 128

RESULT 10
US-08-925-767-7
; Sequence 7, Application US/08925767
; Patent No. 6225084
; GENERAL INFORMATION:
; APPLICANT: FALB, DEAN A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/925,767
; FILING DATE: 09-SEPT-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,573
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/386,844
; FILING DATE: 10-FEB-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-097
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 128 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
US-08-925-767-7

Query Match 1.8%; Score 126.4; DB 4; Length 128;
Best Local Similarity 99.2%; Pred. No. 1.2e-25;
Matches 127; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3120 gggaggtgggctctctatccttaggcacgcacgtgattgttacctgttgccagaaagaa 3179
|||||

Db 1 GGAGGTGGGCTCTGCTCATCTTAGGCATCGCACTGATTGTTACCTGTTTGCAGAAAGAA 60
Qy 3180 taaaatgacataagcaaacatctcttcaaaagtgaggatttccaaatgtcccatatgc 3239
Db 61 TAAAAATGACATAAGCAAACTCATCTTCAAAAGTGGAGATTTCAAAATGTCCCGGTATGC 120
Qy 3240 tgaatacc 3247
Db 121 TGAATACC 128
RESULT 11
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 1.3%; Score 89.8; DB 1; Length 7218;
Best Local Similarity 4.0%; Pred. No. 2.2e-14;
Matches 16; Conservative 253; Mismatches 130; Indels 0; Gaps 0;

Qy 1225 tcgccagttctgcatacatccaaaccttcctctacacaccaccattatgtccgaac 1284
Db 1060 TTGGGATYVV 1119
Qy 1285 actcgggtgtcttgacactgatgtgagttgttagtgactcctcctctctcc 1344
Db 1120 YY 1179

Qy 1345 tctctcctctcttcttcttctcagggcctcttctgctctgctctgtgtcacaatcc 1404
Db 1180 YY 1239
Qy 1405 caccattttttcatcaattttaccatcaaccaggcctctgtgcatctactaaagtct 1464
Db 1240 YY 1299
Qy 1465 acctgatgcataccacacatgctctctccatcacatcacctttaccagtgatccttaacg 1524
Db 1300 YY 1359
Qy 1525 acatctacatctgcccaactttctgtctcacaacaacacttgccacagtcactctctacc 1584
Db 1360 YY 1419
Qy 1585 cctgtcctgccagggcaggagactcctctgtgacttca 1623
Db 1420 YYYYYYYYYYYYYYGTACCAAAATCTTCTATCTCTTTA 1458
RESULT 12
US-08-728-323A-1/c
; Sequence 1, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/728.323A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3489 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3489
US-08-728-323A-1
Query Match 0.8%; Score 56.2; DB 2; Length 3489;
Best Local Similarity 50.2%; Pred. No. 3.2e-05;

Search completed: July 26, 2002, 01:37:45
Job time: 20054 sec

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.
OM nucleic - nucleic search, using sw model
Run on: July 25, 2002, 19:34:16 ; Search time 12075.4 Seconds
(without alignments)
12047.710 Million cell updates/sec

Title: US-09-840-746-2
Perfect score: 6952
Sequence: 1 gttgatgaagaattgcg.....tactgggaagctatagggtg 6952

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_ov.*
5: gb_ov.*
6: gb_ov.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_or.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No. Score Match Length DB ID Description

SUMMARIES

1	5140.4	73.9	5142	9	AB033063	AB033063 Homo sapi
2	5098.4	73.3	6407	6	AR065869	AR065869 Sequence
3	3508.6	50.5	189933	2	AC026342	AC026342 Homo sapi
4	3500.6	50.4	110000	2	AC092983_0	AC092983 Homo sapi
5	1438.8	20.7	1462	9	BC004539	BC004539 Homo sapi
6	1384.8	19.9	1397	9	BC014053	BC014053 Homo sapi
7	1359.8	19.6	110000	2	AC092983_0	AC092983 Homo sapi
8	663	9.5	118407	2	AC026629	AC026629 Homo sapi
9	610.8	8.8	118407	2	AC026629	AC026629 Homo sapi
10	387.4	5.6	400	6	AX245407	AX245407 Sequence
11	326.2	4.7	50921	2	AC092983_3	AC092983 Homo sapi
12	266.4	3.8	115766	2	AC069571	AC069571 Homo sapi
13	263.4	3.8	34308	9	AF389852	AF389852 Homo sapi
14	126.4	1.8	128	6	AR053542	AR053542 Sequence
15	126.4	1.8	128	6	AR080351	AR080351 Sequence
16	126.4	1.8	128	6	AR148306	AR148306 Sequence
17	112	1.6	112	11	G20489	G20489 human STS A
18	102.6	1.5	2008	9	AB062989	AB062989 Macaca fa
19	89.8	1.3	7218	6	I66494	I66494 Sequence 14
20	88	1.3	110000	2	AC092983_2	AC092983 Homo sapi
21	67.6	1.0	150110	2	AC097722	AC097722 Homo sapi
22	67.6	1.0	177475	2	AL626782	AL626782 Mus muscu
23	67	1.0	179262	2	AC079507	AC079507 Mus muscu
24	65.8	0.9	2384	3	PFAGAR	PFAGAR Plasmodium
25	65.6	0.9	397	11	G37798	G37798 GARP Plasm
26	65.2	0.9	56757	2	AC105856	AC105856 Rattus no
27	64.8	0.9	174591	2	AC103093	AC103093 Rattus no
28	64.4	0.9	197844	2	AC079176	AC079176 Homo sapi
29	63.6	0.9	156224	2	AC096059	AC096059 Rattus no
30	63.6	0.9	183272	2	AC094428	AC094428 Rattus no
31	63.4	0.9	91999	2	AC095962	AC095962 Rattus no
32	63.2	0.9	165077	10	AC084382	AC084382 Mus muscu
33	63	0.9	194472	2	AC096438	AC096438 Rattus no
34	62.8	0.9	166704	2	AC078821	AC078821 Homo sapi
35	62.6	0.9	184855	2	AL606744	AL606744 Mus muscu
36	62.6	0.9	209671	10	AC096777	AC096777 Mus muscu
37	62.4	0.9	126736	2	AC044788	AC044788 Homo sapi
38	62.4	0.9	182288	9	AC018362	AC018362 Homo sapi
39	62.4	0.9	203299	2	AP001178	AP001178 Homo sapi
40	62.4	0.9	244490	2	AC105495	AC105495 Rattus no
41	62.2	0.9	82973	2	AC105453	AC105453 Rattus no
42	62.2	0.9	170749	10	AL133159	AL133159 Mouse DNA
43	62	0.9	202904	2	AL603837	AL603837 Mus muscu
44	61.8	0.9	146438	9	HSBA191L9	AL117329 Human DNA
45	61.6	0.9	174990	10	AC027653	AC027653 Mus muscu

ALIGNMENTS

RESULT 1
AB033063
LOCUS
DEFINITION Homo sapiens mRNA for KIAA1237 protein, partial cds.
ACCESSION AB033063
VERSION
KEYWORDS
SOURCE
Homo sapiens brain cDNA to mRNA, clone_lib:pbbluescriptII SK plus
clone:fh09696.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
Nagase,T., Ishikawa,K., Kikuno,R., Hirose,M., Nomura,N. and Ohara,O.

TITLE
Prediction of the coding sequences of unidentified human genes. XV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro

JOURNAL
MEDLINE
DNA Res. 6 (5), 337-345 (1999)

REFERENCE
AUTHORS
Ohara,O., Nagase,T. and Kikuno,R.
Direct Submission

JOURNAL Submitted (04-OCT-1999) Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:cdna@kazusa.or.jp, URL:http://www.kazusa.or.jp/huge/, Tel:+81-438-52-3913, Fax:+81-438-52-3914)

FEATURES Location/Qualifiers

1..5142
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="fh09696"
 /tissue_type="brain"
 /clone_lib="pBluescriptII SK plus"
 1..1966
 /gene="KIAA1237"
 <1..1966
 /gene="KIAA1237"
 /note="Start codon is not identified."
 /codon_start=2
 /product="KIAA1237 protein"
 /protein_id="BA086551.1"
 /db_xref="GI:6330762"
 /translation="SAPLSVSOTTLPQSSSTPVLPRARETPVTSFQTSMTSFTMLH
 SQADADLSQSPHPEKVTESKPSLSLPSTESKAVTINSPLPSLSTESQTLPL
 AFTSNLAQSPFTTILKTSQTLMTTSLSTASLVTPGPIAVQTTACKQUSLTHPE
 ILVPOISTEGGISTERNRVIDATGLIPLTSVPTSAKEMTKLGVTAEYSPASRLG
 TSPQQTWVSTAEADLPKSAFVQSSSTQSPSTLSSASVNSCAVNPCLHNGECVAD
 NTSRGYHCRPSSWQDDCSVDVNECLSNPCSTATCNNTQGSFCKCPGVQLKGI
 CNLVRTEFLKRLFTNLTVEKHSIDLQEVENEITKTLNMCFSALPSVIRSTVNASR
 ESNVAVISLQTTFSLASNVTLFDLADRMQKQVNSKSNVEQLLGSORRIRFASLRC
 KRKSPEDKDSICIDLDGVALCQCKSGTFQFNKMDHSCRACEDGTRENTECMSCPF
 GUGLNGCNXPYGLITWIAAAGGGLLILGIALIVTCRRKNKNDISKLFKSGDFQMS
 PYAEYKPNRSEWREARETEHENGSTKNLQMTDYYVSYPTSVRNPELERNGLYPAYT
 GLPGSRHSCIFPGQYNPSFISDESRRDYF"
 BASE COUNT 1320 a 1317 c 1248 g 1257 t

Query Match 73.9%; Score 5140.4; DB 9; Length 5142;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 5141; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1533 atctgcccacttctgtctcaacaacaccccttgccacagtgatctcttaccctgtcct 1592
 Db 1 ATCTGCCCACTTCTGTCTCAAAAACACCTTGCACAGTCATCTTACCCCTGTCC 60
 QY 1593 gccagggaaggagactctgtacttcatttcagacatcaacaatgacatcattcat 1652
 Db 61 GCCCAGGCAAGGAGACTCTCTGTACTTCATTTCAGACATCAACATGACATTCAT 120
 QY 1653 gacaaatgctcctagtagtcaaaactgcagacaccttaagagccagagcaccacccaaga 1712
 Db 121 GACAATGCTCCATAGTAGTCAAACTGCAGACCTTAAGAGCCAGAGCACCCACCAAGA 180
 QY 1713 gaaagtcaataagaatcaaaagtcacaaagctgggtgtctctgtcccaagagtcacacaa 1772
 Db 181 GAAAGTCATTACAGAAATCAAAAGTCACAAAGCTGGTGTCTCTGCCACAGAGTCCACCAA 240
 QY 1773 agctgtaaacaacaactctcttctgctccatcttaacagagctctccacagacaaac 1832
 Db 241 AGCTGTAAACAAACACTCTCTTTGGCTCCATCTTAAACAGAGTCTTCCACAGAGCAAC 300
 QY 1833 ccttcagcgcacaaagcaccaacttagcacaaaagtctcacaactttcacaaactaccattct 1892
 Db 301 CCTTCCAGCCACAAAGCACCAACTTAGCACAAATGTCTCCAACCTTTCACAACTTACCATTCT 360
 QY 1893 gaagacctcagcctcttattatgaccactcctggcaccctgtcaagcagcagcatctctgg 1952
 Db 361 GAAGACCTCTCAGCCCTCTTATGACCACCTCTCTGGCACCCTGTCAAGCACAGCATCTCTGGT 420
 QY 1953 cactggccctatagcgtacagactcagctggaacacagctctcgtgaccatcctga 2012
 Db 421 CACTGGCCCTATTAGCCGTACAGACTACAGCTGGAAACAGCTCTCGCTGACCCATCCTGA 480

QY 2013 aatactagttctctcaaatctcaacaagaagtggcaccagagaaggaacccagatgat 2072
 Db 481 AATACTAGTTCTCTCAAAATCTCAACAGAAAGGTGCATCAGCACAGAAAGAACCGAGTGAT 540
 QY 2073 tgtggatgtacacactggattgtatccctttgaccagagtgtaccacatcagcaaaaagaat 2132
 Db 541 TGTGGATGCTTACACTTGGATTGTATCCCTTTGACCAGTGTATCCACATCAGCAAAAAGAAAT 600
 QY 2133 gaccacaaagtctggcgtttacagcagagtacagccagagcttcagttccctcggaaacatc 2192
 Db 601 GACCACAAAGCTTGGGCTTACAGCAGAGTACAGCCAGCTTCAGGTTCCCTCGGAACATC 660
 QY 2193 tccttctcccaaaacacacagttgtttccacggtgaagacttggctcccaaaatctgcccac 2252
 Db 661 TCCTTCTCCCAAAACACACAGTTGTTTCCACGGCTGAAGACTTGGCTCCCAAAATCTGCCAC 720
 QY 2253 ctttctgtttcagagcagacacagttccacaaacaaactgtctcttccagcctcagttcaa 2312
 Db 721 CTTTTGTGTTTCAGAGCAGCACACAGTCCACAAACAACTGTCTCTTTCAGCCCTCAGTCAA 780
 QY 2313 cagctgtctgtgaacccctgtcttcacaatggcgaatgcgtcgcagacacaacaccagccg 2372
 Db 781 CAGCTGTGCTGTAACCCCTGTCTTCAATGGGAATGCTCGCAGACAAACACCAGCCG 840
 QY 2373 tggctaccactgcaggtgcccgccttctctggcgaaggagatgattgcagttggtgagtaa 2432
 Db 841 TGGCTACCACCTGCAGGTGCCCGCTTCTTGGCAAGGGGATGATTGCGAGTGGATGTGAA 900
 QY 2433 tgagtgccctgtgaacccctgtccatccacagccagtcgaacaataactcagggatcctt 2492
 Db 901 TGAGTGCTGTGCAACCCCTGCCCATCCACAGCCACGTCACAAATACTCAGGGATCCTT 960
 QY 2493 tatctgcaaatgcccgggtgtgggtaccagttggaaaaggagatatgcaatttggtagaac 2552
 Db 961 TATCTGCAAAATGCCGGTGGGTACCAGTTGGAAAGGGATATGCAATTTGTTAGAAC 1020
 QY 2553 ctctgtgacagagtttaaatataagagaactttctttaaataacactgtggaacaaatc 2612
 Db 1021 CTTCGTGACAGAGTTTAAATATAAGAGAACTTTTCTTAATACAACTGTGAAAACATTC 1080
 QY 2613 agacctacaagaagtggaaaatgagatcaccaaaaagttaaatatgtgtttttcagcggt 2672
 Db 1081 AGACCTACAGAAGAGTTGAAAATGAGATCACCAAAACGTTAAATATGTGTTTTTCAGCGGT 1140
 QY 2673 acctagtacatccgctctacagttcacgctctgaggagtcacaaacgctgggtgagtcctc 2732
 Db 1141 ACCTAGTTACATCCGATCTACAGTTACGCTCTAGGAGTCCACACGCGTGGTGTATCTC 1200
 QY 2733 actgcaacaacacctttccctggcctcccaatgtgacgctatttgacocctggtgataggat 2792
 Db 1201 ACTGCAAAACAACTTTTCCCTGGCTCCAATGTGACGCTATTTGACCTGGCTGATAGGAT 1260
 QY 2793 gaagaatgtgtcaactcctgaagtcctctgtgaggtcgtgagctcctgtggatctcgaatcga 2852
 Db 1261 GCAGAAATGTGTCAACTCTCTGCAAGTCCCTGTGTGAGGTCTGCGAGCTTTGGGATCTCA 1320
 QY 2853 gagcgagatcttagagcgggcagcttctgcaagcgaagaagtcgccgaatgtgacaaaaga 2912
 Db 1321 GAGCGGATCTTTAGAGCGGCGAGCTTGTGCAAGCGGAAGAGTCCCGAATGTGACAAAGA 1380
 QY 2913 caactccatctgcactgacctggacggcgttggcctgtgacctgtgcaagtcgggatactt 2972
 Db 1381 CACCTCATCTGCACATGACCTGGACGCGTTCCTGTGCGAGTGGCAAGTCGGGATACTT 1440
 QY 2973 tcaagtcaacaagaatggaccactcctgcgagcatgtgaagatggatataaggcttgaaga 3032
 Db 1441 TCAGTTCAACAAGATGACCACTCTCCGAGCATGTGAAGATGGATATAGGCTTGAAAA 1500
 QY 3033 tgaacctgcatgagttgcccatttggccttgggtctcaactgtggaacccccctatca 3092
 Db 1501 TGAACCTGCATGAGTTGCCCATTTGGCTTGGTGTCTCACTGTGGAACCCCTATCA 1560
 QY 3093 gcttataactgtgtgtagtcgacggcgggaggtggggtcctcgtctcatctcctagggcatcgc 3152

|||||
Db 1561 GCTTATCACTGTGATGCGAGCGCGGAGGTGGGCTCTGCTCATCTTAGGCATCGC 1620
Qy 3153 actgattgttacctgttgcagaagaataaaaaatgacataagcaaaactcatcttcaaaag 3212
Db 1621 ACTGATTGTTACCTGTGTCAGAAAGAAATAAAATGACATAAGCAAAATCATCTTCAAAAG 1680
Qy 3213 tggagatttccaaatgccccatagtgaatacccccaaaaaactcctgcgtcacaaagt 3272
Db 1681 TGGAGATTTCAAATGTCCCGGTATGCTGAATACCCCAAAATCTCTCGCTCACAAAGATG 1740
Qy 3273 gggccgagaagctattgaaatgatgagaatggaagtaccaaaaaacctctccagatgac 3332
Db 1741 GGGCCGAGAAAGCTATTCAAATGATGAGATGGAAGTACCACAAAACCTCTCTCCAGATGAC 1800
Qy 3333 ggatgtactactcgctcacaaagtgttaagaaatccagaacttgaaacgaacgactcta 3392
Db 1801 GGATGTGTACTACGCTTCAAGTGTAAAGTGTAAAGAACTTGAACGAACGACACTTA 1860
Qy 3393 ccggtctacactggactgcaggatcacggcattcttgcatttcccgagacagtataa 3452
Db 1861 CCGGCCCTACACTGGAGCTGCCAGGATCAGCGCATCTTGTCATTTTCCCGGACAGTATA 1920
Qy 3453 ccggtcttcatcagtgatgaaagcagaagaagagactacttttaagtcacagagagaga 3512
Db 1921 CCGTCTTTTCATCAGTCATGAAGCAGACAGAGAGACTACTTTAAGTCCAGGAGAGAGA 1980
Qy 3513 ggaactcattgctgagcagtcacactgggaactctgctcagagaccgacacacagagag 3572
Db 1981 GGGACTCATTTGCTGTAGCCAGCTACCTCGGACCTCTGCTCAGAGGACCCACAGAGG 2040
Qy 3573 ctgcgccagagatttgcggagccagctgagtggaagcaggaagaagagagcagcatg 3632
Db 2041 CTGCGCCAGAGATTGTCGGAGCCACGCTGAGTGGCAAGCAGGAGGACAGGCATG 2100
Qy 3633 cggggcgtagccacagtgaggagagacaggtggatgtggaaccacagcgtcctcattcagc 3692
Db 2101 CCGGGCTGACACAGCTGAGGAGACAGGTGGATGTGGAACACACAGGCTGCTCATTCAGC 2160
Qy 3693 acctgtgttactgtgaacgtgaatgtggccagatatacaagagagctctctgagtga 3752
Db 2161 ACCTTTGTGTACTGTGAACGTGAATGTGGCCAGTATCAAGAGAGTCTCTCTGAGTGA 2220
Qy 3753 ctgcacacatggcactggcaccagggcgactattaggcaggcgacacactagacttcagt 3812
Db 2221 CTGCACCATGGCACTGACACCGGCGACTATTAGCCAGGGCAGACCACTAGACTTCAGT 2280
Qy 3813 gcagggaacctgggtttccctctogtttgcacttttagtaaaatggggtggaggtttcccttt 3872
Db 2281 GCAGGGACCTGGTTTTCCCTTCGTTTTCGACTTTAGTAAATGGGTGGGAGGTTTCCCTTT 2340
Qy 3873 ggaactgtttgagactgttccagaagaaggcttcccttcccgagacacacttccataggc 3932
Db 2341 GGATCTGTTTGTAGACTGTCTCCAGAAAGAGGCTTCCTTTCCCGAGACACTTCATAGGC 2400
Qy 3933 agcaattgtgtattcttcgagcaaaaactggctgttgaattatttctcgtgccagc 3992
Db 2401 AGCAATTTGGTGATTCTTTCAGCAAAATACTGGCTGTGTTAATTTCTTCTGCCAGC 2460
Qy 3993 gctgcgtgtaaacacacagatgagatgagcgtaccactgaagctgaaagtgtgcacca 4052
Db 2461 GCCTGCGTGTAAACACACATGAGGATGAGCGTACCACCTGAAGTCTGAAGATGTCGCCA 2520
Qy 4053 ttgaacggacagtggttttctatagtcttctaggttgtctttagctacagtttccaagccag 4112
Db 2521 TTGAACGGACAGTGTTTTCATATGTCTTCTAGTGTGCTTATGCTACAGTTTCCAAGCCAG 2580
Qy 4113 cccccagtgaggaagtgtgagccacgcacacaactgcaatgtgttttttaagtca 4172
Db 2581 CCCCCAGTGAGGAATGTGTAGGCACCGCACACAACCTGCAATGTGTTTTTAAGTCA 2640
Qy 4173 aggtgacacatgtatttaagattttttttaaatactcttgcagtttaaatctcacctttt 4232
|||||

Db 2641 AGGTGACACATGTATTAAAGATTTTTTTTTTAAAAATCTCTTTGTCAGTTAAATCTCACTTTT 2700
Qy 4233 tcaacaagcctggatgcagggcacaacaacttatatttggtttagctggaggctcagca 4292
Db 2701 TCAAAACAGCCTGGATCAGGGCAAAACAACTTATATTGGTTTTAGCTGGAGGCTCAGCA 2760
Qy 4293 ggcagattgcaggcaggggggaccttttccatgcagggcccaagcctgggctgggac 4352
Db 2761 GGCAGATTGAGGAGGGGGGACATTTTCATPCCATGAGGGCCAGCCTTGGGGCCTGGGAC 2820
Qy 4353 tctgataccactgtgagggccagagcagctgcgtatgagggagaaatgtccaactgaa 4412
Db 2821 TCTGATCACCATTTGTGGAGCCAGAGCAGCTGCTGATGGAGGAGAAATGTCAAACGTAA 2880
Qy 4413 cgaagtttccacactctagaaagcagctgtttagccccctgcagctggaatggttag 4472
Db 2881 CGCAGGTTTCACTACTAGAAAAGCAGCTTGTGAGCCCCCTGCAGCTGGATGTGGTTAG 2940
Qy 4473 agggatggctgaatagcaggttagatttccctgcatacaacagctgcttgggaagctgtg 4532
Db 2941 AGGATGGCTGAATAGGCAGGTTAGATTTCTCTGCATCAACAGTGCCTTTGGGAAGCTGTG 3000
Qy 4533 tggattcctgaggaagacacaggggagcggatggagccacacatgagtttgcctcacggc 4592
Db 3001 TGGATTCTGAGGAAGAACAGGGAGCCGAGATGGAGCCACACATGAGTTTGTCTCACCGGC 3060
Qy 4593 tactgcagcattgttaccagaaatctcatgtccaaaaacccatgtaaaactttcaacca 4652
Db 3061 TACTGTCAGCATTGTGTACCAGAACTCATGTCCACAAAACCCCATGTAAACTTTCAACCA 3120
Qy 4653 ctcaaacgttattcgcgtggaagaataaacttttttctcaccagctcattgttacc 4712
Db 3121 CTCAACGCTGTTATTTCGGCTGAAGAAATAACTTTTTTTTCTCACCCAGTCATTTGTACC 3180
Qy 4713 tcttcattggctgtgtgcacccctccagaaacgtggttatacttccagtcagtggtgga 4772
Db 3181 TCTTCATATGCTGTGTGCGACCCCTCCAGAAACGTGGTTATATTCTCAGTCAGTGTGGGA 3240
Qy 4773 gaactgaagacttccggttggctgaggaactgaggttgacccctcgggaaggaagtcca 4832
Db 3241 GAACGAAGACTTCCGGTTGGTCGAGGAACGTGAGGGTTGACCTTCGGGAAGGAAGTTCCA 3300
Qy 4833 ctcatctatttattatgtcgtgtgagtcctgcagggagacatccagtcactcgg 4892
Db 3301 CTCATCTTATTATTATGCGCTGTGATGTGGTCTCGCCAGGGAGACATCCAGTACTCGGT 3360
Qy 4893 gtctttaatgtccacctggggaaactgtgttatttggccttcttggggcatcctcgggttt 4952
Db 3361 GTCTTTTAATTGCCACCTTGGGGAACGTGTATTATTGGCTTCTTTTGGGCATCTCTGGTTT 3420
Qy 4953 ggaatgaagtgaaggaatacagaggttaaaagaattgtccacccctgaagcgggagctc 5012
Db 3421 GGATGAAGTGAAGGGAATACAGAGTAAAGAAATGTGTCCACCCCTGAAGCGGGAGTCC 3480
Qy 5013 cgcttcacattctcggaaatgggtgcagccactggggagcagttctgcgccggcagtggtg 5072
Db 3481 CGCTTCACATTTCTGGAATGGTGCAGCCACTGGGACAGTTCGCCCGGGCATGGTTG 3540
Qy 5073 ttcttcaaggtcctcctaaataataatccctatttctacataatccttggcctgatggtt 5132
Db 3541 TTTCTTTCAAGGCTCTCTAAATATAATCCCTATTCTTACATAATCCTTTGGCCCTGATGGTT 3600
Qy 5133 ttaagaagaactcctgtgtcccatggtctccaccaactcaccatcaacctgtctgtagcaa 5192
Db 3601 TTAAGCAAGAACTCTCTGTGCCATGGTCTCCACACTCCACATCACCCCTGCTGTAGCAA 3660
Qy 5193 gactcctagtcagggaggtgcatttttagtttagttaattgcacttaccatgagataaat 5252
Db 3661 GAGTCTCTAGTCAGGGAGGTGCATTTTAGTAGTTAAATTTGCATTTATCCATGAGATAAT 3720
Qy 5253 aaaaaggagaactgttttttatcagtgagggttaacctaaaaatttcaaatgtgcgctttt 5312
Db 3721 AAAAGGAGAACTGTTTTTATCAGTGGAGGCTAACCTTAAATTTCAAAAGTGTGCGCCTTTT 3780

Qy	5313	gaaatcttgggcctctctctctgtagaaccaattggcccttttgtggtcagcgccctcgac	5372
Db	3781	GAAATCTTGGGCCTCTCTCTGTAGAACCAATGSCCCTTTTGCGCTCACGGCTCGCAC	3840
Qy	5373	ctaactggagagttcttgaaactctgacgtcacgtcacctgaaccacacagactaggcttcctggc	5432
Db	3841	CTAACTCGAGAGTTCGTAGCTCTCGAGCTACACTGAGCCACACAGACTAGGCTTC TTGGC	3900
Qy	5433	tccttcgcgacgatgcctgctcaccccagaaccgcgagctgtgggaagagccaatgtagg	5492
Db	3901	T CCTTCGCAGCATGCTGCTCAOCCCCAGAACCCGCGAGCTGTGGGAGAGCCATGTAGG	3960
Qy	5493	gaggctattccccaggcatacaattocactgccttcagctgacgtcagctcagctgcacaaatca	5552
Db	3961	GAGGCTATTCCCAGGCATACATTCCACTGCCCTTCAGCTGACGTCACAGCTCAGACAAATCA	4020
Qy	5553	tctcctctatcggagccagaagaacttcagctccacaaaatgaagtgtctgtcctcgaaaa	5612
Db	4021	TC TCCTCTATTCGGAGGCGAAGAATTCAGCTCCACAATAAGAGTGTCTGCTCTGAAAA	4080
Qy	5613	cattcttgggaagaaatcccaaacatcgagaaaaacggtctcctgtgagttcccaacaatgctt	5672
Db	4081	CATTCTTGGGAAGAAATCCCAACATCGAGAAAACGGTGTCTGTGNGTTCCAACAATGCTT	4140
Qy	5673	cttgttcatgggtttbtctcgtatggatggattaagagtgctttattttgttgtcttaa	5732
Db	4141	CTTGTT CATGGGTTTCTCCGTATGGAGTGGATT AAGAGTGT TTA TTTGTTCTTCTTAA	4200
Qy	5733	ctgagaaaaaaggagccaccaacagaagttgaggtcacacagttctccacagttlccaggga	5792
Db	4201	CTGAGAAAAAAGGAGACCACCCACAAAGTTGAGGTTCACACAGTCTCCACAGTTTTCCAAGGA	4260
Qy	5793	ggqgtttgggggtggggaaaggcacctccagagacatgaggtcctaaggggacatgataa	5852
Db	4261	GGCGTTTGGGGTGGGAAAGGCACCTCCAGACCATGAGGCTCTAAGGGGACATGAGTAAA	4320
Qy	5853	gcatgtctgtgacccagtgagaaaggagggccagctgcactcctcctgcacoggggtlcccta	5912
Db	4321	GCATGTCTGTGACCCAGT GAGGAAGGAGAGGCCAGCTGC ACTCTCTGCACGGGCTTCCTA	4380
Qy	5913	gctgcagaaggtlcccgcttagccgaggggaaacacctgatacagaaagaagcctgagat	5972
Db	4381	GCTGCAGAAAGGTTCCCGCTTAGCCGAGGGGAAACACCTGATAGCAGAAGAGGCTCGAT	4440
Qy	5973	gcacacctggcacgcgaggtctctgcgccagacagatgtctccatgtcagcccccgcac	6032
Db	4441	GCACACCTGGCACGCGAGGCTCTCCGCCACAGACACATGCTCCATGT CAGCCCCCTGCAC	4500
Qy	6033	ctgggggtgtgtattcaagtgcacagatgcacaaatcctgcaccaaatatccccacagatgg	6092
Db	4501	CTGGGGTGTGTGATTACGTGTGCACAGATGCCACAAATCCTGCACCANAATCCCCACAGATGG	4560
Qy	6093	gggaaggtgagagaaaggcgcaagtgaatgttaactgctcaagagatgctttaaacctcca	6152
Db	4561	GGGAAGGTGAGAGAAAGGGCAAGTGATGTGTA CTCTCAAGAGATGCTTTAAACCTCCA	4620
Qy	6153	tagagaggagccggggcgagggggcatctgtgtccccgtcacacactgcagcaggggaagg	6212
Db	4621	TAGAGAGAGCCGGGCGCAGGGGSCATCTGTGTGTCCCCGT CACACACTGCAGCAGGGAAGG	4680
Qy	6213	gtggctggctggctccctggatcagttggttttggtttaagctccagagggtcttatggcc	6272
Db	4681	GTGGCTGGCTGGCTCCCTGGCATAGTGGTGT TGGTTTAAGTCTCAGAGGGTCTTATTGGCC	4740
Qy	6273	attgtcttttctctgccccttgagccagcctaaaggccctggagtcgtgttcttctttagcg	6332
Db	4741	ATTGTCTTTTCTCTCGCCCTTGAGCCAGCCTAAGGCCCTTGAGTCTGTCTTTTAGGGG	4800
Qy	6333	gatgaactgacatgctctctaccatgaccaggctctggggcagaaggctcctcacagtatcctt	6392
Db	4801	GATGAAC TGACATGCTCTTACCATGACAGGCTCTGGCAAAGGCTCTCAGAGTATCCTT	4860

RESULT 2

AR065869

LOCUS

DEFINITION

ACCESSION

VERSION
KEYWORDS

KEYWORDS
SOURCE

SOURCE
ORGANIZATION

Findings

REFERENCES

AUTHOR:

TITLE

ANNALS

JOURNAL FEATURES

FEATURES

302

BASE COU

ORIGIN

0123456789

Query
Best

Query Match 73.3%; Score 5098.4; DB 6; Length 6407;
Best Local Similarity 94.3%; Pred. No. 0;
Matches 5412; Conservative 21; Mismatches 8; Indels 301; Gaps 2;

•

Db 1026 TGATGGTGGCCGAGAACGCTGCGATCTTTGACGGTCAGTCTGGGACCTGTGACGAAGAC 1085
Qy 421 agaaggcttccccagactccagaattgccaagacttcatctcagctcttcttccacc 480
Db 1086 AGAAGGCTTCCCCAAGACTCCAGAAATGGCAGGACTTCATCCTCAGTCTCTTTTCCACC 1145
Qy 481 ctctcgagtggaatcgagaagaacagtagagtaactgggaatccacaggggatgaggaatt 540
Db 1146 CTCTGCACTGGATCGAGAAGAACAAGTAGAGTAACCTGGGAATCCAGGGGATGAGGAATT 1205
Qy 541 cattgaaccatccacagaaataaatttggacttaactcttcttgcg-tggcaaaatgattc 599
Db 1206 CAUTGAACCATCCACAGAAATGAATTTGGACTTACGCTTTGGGTTGGCAAAATGATTC 1265
Qy 600 cccaaccttggagaacatcagcttgcagcagctctcagagtgcaaaatggaagtcocat 659
Db 1266 CCCAACCTTTGGAGAACATCAGCTTGCAGCAGCTTCGAGGTGCAAAATGGAATGCCAT 1325
Qy 660 gtctcagactgagactgtctaggtcagtcgcacccatgagagtgagagatcactgc 719
Db 1326 GTCTCAGACTGAGACTGTGTCTAGGTGAGTCGCACCCATGAGAGGTGAGAGATCACTGC 1385
Qy 720 aactggcttggaccacagcaacacatctgcagatgtgacagggagctctgtctcata 779
Db 1386 ACATGGCTCTTGACCAACAGCACAAACATCTGCAGATGTGCAGGAAGCTCTGCTTCATA 1445
Qy 780 tectgaagtgtaagtctcagtggttgcaccagttctcagactctactgtacagctgg 839
Db 1446 TCCTGAAGGTGTGAATGCTTCAGTGTGACCCAGTTCAGACTCTACTGTACAGTCTGG 1505
Qy 840 aggaagtcacacagcattggagatagaggttattcagagctcttcacatctctc 899
Db 1506 AGGAAGTCACACAGCATTGGAGATAGGAGTTATTACAGACTTTCATCTACATCTTCTC 1565
Qy 900 ggaagcttgaattcatcagcaccacgaogtggagaacttcaa----- 940
Db 1566 GGAAGCTTGTAATTCATCAGCACCACGCTGGAGAACGTTCAACCTTGGAGACAGCCGAGA 1625
Qy 941 ----- 940
Db 1626 GCCAGCCCAAGCACTAGGTGACAGTTCGCCCAATGACAGGAGCAGGACTTCTGGGGTGCC 1685
Qy 941 ----- 940
Db 1686 CTCTCTCGGCACCCACACCTTGGCTACTGTCACTTGGAAACGGGGAGCAGCAGCTGCGGTC 1745
Qy 941 ----- 940
Db 1746 TGTACCCCTCACCAACACCAGCATGAGCAGGACTTCTGGGGAGCAGGCGCTGCGAGC 1805
Qy 941 ----- 940
Db 1806 GGCCATGCCCAAGAAACAGAGGGTGCTCTCTGACAGTAAACGTGACGGACGACATGGG 1865
Qy 941 -----tcgtgggattagctacgg 959
Db 1866 CTTGGTCTCAGGTCAGTGGCCGCTCCAGTCGACTCGGAGTCTGGGGATTAGCTACGG 1925
Qy 960 tcaagtgcgtggcacagctattgaacaaaggacttccagcgaccacacagaccacta 1019
Db 1926 TCAAGTGCCTGGCACAGCTATTGAACAAAGGACTTCCAGGCGCCACACAGACCACCTA 1985
Qy 1020 cctgtcatctacttccacaaaggagaacggcggttactgtccattacagataacagttc 1079
Db 1986 CTTGTATCTACTTTCACCAAGAGGAGAACGGGGTTACTGTCCATTACAGATAACAGTTC 2045
Qy 1080 atcctcagacattgtgagagctcaactctctatatataaaatctcaaaccttccacttc 1139
Db 2046 ATCCTCAGACATTTGGAGAGCTCAACTTCTTATATTAATAATCTCAAACTCTTCACATTC 2105
Qy 1140 agagtattctctcttctctcagactgagagaagtaacatctcactatgacgg 1199
Db 2106 AGAGTATCTCTCTTTTCTCATGCTCAGACTGAGAGAGTAACATCTCATCTCATGACGG 2165

Qy 1200 ggaatatgtcagcctctactagtcgacagttctctgcatatcatccaaaccttccgtctcta 1259
Db 2166 GGAATATGCTCAGCCTTCTACTAGTGCAGGTTCTGCATACATCAACACCTTCGTCCTA 2225
Qy 1260 caeccccaccatataatgcgaacacttgggttgttcttgacactgatgctgagttgt 1319
Db 2226 CACACCCACCATTAATATGCGAACACTTCGGTGTGTTCTGGACACTGATGCTGAGTTGT 2285
Qy 1320 tagtgcct 1379
Db 2286 TAGTGACT 2345
Qy 1380 gcctctgcctctgtgcacaatcccaacttatttccatcaattttacatcaaacag 1439
Db 2346 GCCTCTGCCCTCTGTGTGCAATCCCACTTTATTTCATCAATTTTACCATCAACAG 2405
Qy 1440 ggctctgtgcattactactaaagtctactctgtgcacccacacatgggttctctcacc 1499
Db 2406 GGCTCTGTGCATCTACTAAGTCTACTCTGTGATGCCACACATGGTCTCTCTCACC 2465
Qy 1500 atcaacttaccagttactttaaagacatctacatctgccccacttctctctcacaac 1559
Db 2466 ATCACCCTTTACCAAGTATCTTAAGACATCTACATCTGCCCACTTCTGTCTCACAAC 2525
Qy 1560 aacttgcacagtcattcttaccctgtcctgccagggcagggagagactcctgtgac 1619
Db 2526 AACCTTCCCAAGTCACTTCTTACCCTGTCTGCCAGGGCAAGGGAGACTCTCTGTGAC 2585
Qy 1620 ttcatcttcagacatcaacaatgacatctcatgacaaatgctccatagtagtcaaacgc 1679
Db 2586 TTCAATTTACAGACATCAACAATGACATCATTCATGACAATGCTCCATAGTAGTCAAACTGC 2645
Qy 1680 agaccttaagcagcagcagcccccacacacagaaagtcttattacagaatacaaaagtcacc 1739
Db 2646 AGACCTTTAAGAGCCAGAGCAGCAGCAGCAGCAAACTTCCAGGCCACAAGCAACTTAGC 2705
Qy 1740 aagctgtgtctctgcacacagtcacacaaagctgttaacacaaactctcttggc 1799
Db 2706 AAGCTGTGTCTCTGCCACAGAGTCCACAAAGCTGTAAACAACAACTCTCTCTTTGCC 2765
Qy 1800 tccatcttcaacagagtcctccacagacaaaccttccagccacacagcagcaacttagc 1859
Db 2766 TCCATCTTTAACAGAGTCTCCACAGAGCAAACTTCCAGGCCACAAGCAACTTAGC 2825
Qy 1860 acaaatgtctccaaacttccaaactccatcttgaagaactctcagctctcttcttgacac 1919
Db 2826 ACAAAATGCTCCAACTTTCACAACTTCTGAAGACCTCTCAGGCTCTCTTATGACAC 2885
Qy 1920 tectggcaccctgtcaagcagacatctctgttcaactggccctatagccgtacagactac 1979
Db 2886 TCCTGGCACCCTGTCAAGCAGCAGCATCTCTGGTCACCTGGGCCCTATAGCCGTACAGACTAC 2945
Qy 1980 agctggaaaaacagctctcgtgaccttccctgaaataactagttctctcaaatcttcaacaga 2039
Db 2946 AGCTGGAAAACAGCTCTCGCTGACCCATCTGAAATACTAGTTCCTCAATCTCAACAGA 3005
Qy 2040 aggtggcatcagcacagaaaggaaacccagtgattgttgatgtctaccactggattgacctcc 2099
Db 3006 AGGTGGCATCAGCACAGAAAGAACCGAGTATTGTGGATGCTACCACTGGATTGATGCC 3065
Qy 2100 tttagaccagtgtaaccacatcagcaaaagaaatgacacaaaagcttggcttaacagaga 2159
Db 3066 TTTGACCAGTGTACCCACATCAGCAAAAGAAATGACCACAAAGCTTGGCGTTACAGCAGA 3125
Qy 2160 gtacagcccagcttcagttccctcggaacatctctctctccccaaaccacagttgttc 2219
Db 3126 GTACAGCCCAAGCTTCAGTTCCCTCGGAACATCTCCTTCTCCCCAAACACAGTTGTTTC 3185
Qy 2220 cagggctgaagacttggctcccaaatctgcacacttctgtgttcagagcagcacacagtc 2279
Db 3186 CACGGCTGAAGACTTGGCTCCCAAAATCTGCCACTTGTGCTTTCAGAGCAGCAGCACAGTC 3245

Db 5406 GCTTGTGACCCCTGACACCTGATGCTGTAGAGGATGGCTGAATAGSCAGGTTAGA 5465
 QY 4500 ttctcctcatcaacagctgtttggaagctgtgtgattcctcaggaagaacagggagcc 4559
 Db 5466 TTTCTCTCATCAACAGTGTCTTGGGAASCTGTGTGATTCTCTGAGGAAGAACAGGGAGCC 5525
 QY 4560 gagatgagccacacatgagtttgcctacccggtctactgcagcactttgtaccacagaatct 4619
 Db 5526 GAGATGGAGCCACACATGAATTGCTCACCGGCTACTGCGAGCACTTGTACCCAGAATCT 5585
 QY 4620 catgtccacaaacccatgtaaacctttcaaccactcaaaagctgtttattcgttgaaaga 4679
 Db 5586 CATGTCCACAACCCCATGTAAACTTTCAACACCACTCAAGAGTGTATTTCGGTGAAGAA 5645
 QY 4680 ataactttttttctaccacagctattgtacctcttcatatgctgtgtcgaacctcc 4739
 Db 5646 ATAACTTTTCTTCTACCCAGTCAATTTGACCTCTTCATATGGSTATGTGCGACCCCTCC 5705
 QY 4740 agaaacgtggttactctccagtcagtggtggagaactgaagactlccggttggctgag 4799
 Db 5706 AGAAGCTGGTTACTKCCAGTCAGTGTGGGAGACTGAAGACTTCCGGTTGGTCGAGG 5765
 QY 4800 aactgaggggtgaccttcgggaagtgctccactcactcatttatttatttgcctgtgat 4859
 Db 5766 AACTGAGGGTTGACTTCGGGAAGGAGTTCACACTCATCTATTATTATCCCTGTGATG 5825
 QY 4860 tgggtctccagggagacatcagctactcagtggtgtctttaaattgcacctggggaactgt 4919
 Db 5826 TGGGTCTGCGAGGAGACATCCAGTACTCGGTGTCTKTAATTCGCCACCTGGGGAAGTGT 5885
 QY 4920 gttattggcctcttttggggcctcctggtttggatgaagtgaaggggaatacacagagta 4979
 Db 5886 GTTTATTGGCTCTTTGGGGCATCTCGKTTGGATGAAGTGAAGGGAATACAGAGGTA 5945
 QY 4980 aagaattgtctccacccctgaagcggggagtcctcgttcaacatttcttggaatggtgcag 5039
 Db 5946 AAAGAATTTGCTCACCTCAAGCGGGAGTCCCGGTTTCACATTTCTGGAATGTTGCAG 6005
 QY 5040 ccactgggaacagttctgccccgggcatggtgtttcttctcaaggtcctcctaaataaact 5099
 Db 6006 CCACTGGGGACAGTTCTGCCCCGGGATGTTGTTCTTCAAGGTCTCTAAATATAATC 6065
 QY 5100 cctattctacataactcttggccctgagtggttttaagcaagaactcctgtgccccatgg 5159
 Db 6066 CCTATTCTTACATAATCTTGGCCCTGATGTTTAAAGCAAGAACTCTGTGTCMAATGG 6125
 QY 5160 tctccaccactcaccatcaccctcgtctgtagcaagagtccttagtcaggggaggtgcatttt 5219
 Db 6126 TCTCCACCACTCAACCATCACTCTGCTGTAGCAAGAGTCTCTAGTCAGGGGAGTGCATTT 6185
 QY 5220 agtagtaattgacattatccatgagataataaaaggagaactgtttttatcagtga 5279
 Db 6186 AGTAGTTACATTTGCATTTATCCATGAGATAAATAAAGGAGAGTCTGTTTTATCAGTGA 6245
 QY 5280 ggctaactaaatttcaaaagtgcctcttttgaactcttggcctctctctctctgtaga 5339
 Db 6246 GGCTAACCTAAATTTCAAGTGTGCTTTTGAATCTTGGGCTCTCTCTCTGTAGA 6305
 QY 5340 acaactggcccttctggtcagcggcctcgcacctaactcagagagtgctcagctcgtca 5399
 Db 6306 ACCAATGCCCTTTGTGGTTCAGCGGCTCGCACCTAACTGGAGAGTTCGAGCTCTCGCA 6365
 QY 5400 gctcactgagccacacagactaggtcttcttggtcctctccgc 5441
 Db 6366 GCTCACTGAGCCACAGACTAGGCTTCTTGGCTCTCTCCGC 6407

RESULT 3
 AC026342/c AC026342 189933 bp DNA linear HTG 31-JAN-2002
 DEFINITION Homo sapiens chromosome 3 clone RP11-559L12, WORKING DRAFT
 SEQUENCE, 7 unordered pieces.
 ACCESSION AC026342

VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

COMMENT

AC026342.21 GI:18449521
 HTG: HTGS_PHASE1; HTGS_DRAFT.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 189933)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,
 Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
 Barbaria,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhaq,C., Burch,P., Burrell,C., Burrell,K.L., Byrd,N.C.,
 Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B.,
 Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,D.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mayhew,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
 Moser,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,N., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,S.,
 Nguyen,N., Nickerson,E., Nwokoko,S., Ogih,M., Okwuonu,G.,
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
 Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
 Scherer,S., Scott,G., Shen,H., Shooshkari,N., Sisson,I.,
 Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmani,K., Vazquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G., and Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 189933)
 Worley,K.C.
 Direct Submission
 Submitted (22-MAR-2000) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jan 31, 2002 this sequence version replaced gi:15625920.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: HARC
 Center clone name: RP11-559L12
 ----- Summary Statistics
 Sequencing vector: M13: L08821
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap; version 0.990329First call to
 findPhrapList
 Consensus quality: 189260 bases at least Q40
 Consensus quality: 191512 bases at least Q30
 Consensus quality: 192994 bases at least Q20

Estimated insert size: 189583; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-fp estimation
 Quality coverage: 6.1x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 82312: contig of 82312 bp in length
 * 82313 82412: gap of unknown length
 * 82413 117485: contig of 35073 bp in length
 * 117486 117585: gap of unknown length
 * 117586 147623: contig of 30038 bp in length
 * 147624 147724: gap of unknown length
 * 147724 166827: contig of 19104 bp in length
 * 166828 166927: gap of unknown length
 * 166928 179740: contig of 12813 bp in length
 * 179741 179840: gap of unknown length
 * 179841 186370: contig of 6530 bp in length
 * 186371 186470: gap of unknown length
 * 186471 189933: contig of 3463 bp in length.

FEATURES
 source

1. 189933
 /organism:"Homo sapiens"
 /db_xref:"taxon:9606"
 /chromosome:"3"
 /clone:"RP11-559L12"

BASE COUNT 54758 a 42732 c 41678 g 50156 t 609 others
 ORIGIN

Query Match 50.5%; Score 3508.6; DB 2; Length 189933;
 Best Local Similarity 99.4%; Pred. No. 0;
 Matches 3553; Conservative 0; Mismatches 19; Indels 3; Gaps 3;
 QY 3342 ctactgcctacagtgtaaggatccagaaacttgaaacaggactctacccggccta 3401
 Db 155141 CAACCGCCCTACAAAGTGAAGGAATCCAGAACTTGAAACGAACGGACTCTACCCGGCCCTA 155082
 QY 3402 cactggactccaggatcacggcattcttcgattttcccgagacagataaacccgtcttt 3461
 Db 155081 CACTGGACTCCAGGATCAGCGCATCTTGCATTTTCCCGGACAGTATTAACCCGCTTT 155022
 QY 3462 catcagtga tgaagcagaagaagagactacttttaagtccagagagagagactcat 3521
 Db 155021 CATCAGTGTATGAAGCAGAAAGAGAGACTACTTTTAAGTCCAGGAGAGAGGGACTCAT 154962
 QY 3522 tgccttgagccagctaccttgagacctctgctcagagaccgacccaggaggtcgccca 3581
 Db 154961 TGCCTTGAGCCAGTCACTGGGACCTCTGCTCAGAGGACCGCACCAGGAGGCTGCGGCCA 154902
 QY 3582 ggattgtcgggagccagctgagtgagcagcagagagagagagagcagcagcgggcgctg 3641
 Db 154901 GGATTTGTCGGAGCCACCGCTGAGTGGCAGCAGGAGAGGAGGATCGCGGGCGTG 154842
 QY 3642 accacagtgagagagacagtgatgtggaaccacacaggtctctctcagcagcactttggt 3701
 Db 154841 ACCACAGTGAGGAGACAGGTGGATGTGGAAACACACAGGCTGCTCATTCAGCACCTTTGTT 154782
 QY 3702 gttactgtgaactgaatgtggccagatcatcaagagagaggtctctctcagcagcaccat 3761
 Db 154781 GTTACTGTGAACGTGAATGTGGCCCAAGTATCAAGAGAGTCTCTCTGAGTGACTGCACCAT 154722
 QY 3762 ggcactggcaccagcgagctattagccaggcgagaccactagactcagtcagcgagacc 3821
 Db 154721 GGCACGTGGCACCAGGGCGACTATTAGCCAGGGCAGACCACCTAGACTTCAGTGCAGGGACC 154662

QY 3822 tggttttccctctgcttgcacttttagtaaatgggtggagggtttctcttttggatctgtt 3881
 Db 154661 TGGTTTCCCTTCGTTGTCACCTTAGTAANTGGTGGGAGGTTTCCTTTTGGATCTGTT 154602
 QY 3882 ttgagactgtccagaaaaaagaagcttctcttcccgagacacacttccatagcagcaatttg 3941
 Db 154601 TTGAGACTGTTCAGAAAGAAGGCTTCTTTCCCGAGACACTTTCATATAGCAGCAATTTG 154542
 QY 3942 gtattcatttcagcaaaaataactgctgttatttatttcttcccgagcgtcggtg 4001
 Db 154541 GTATTCTATTGTCAGCAAAATACTGGCTTGTTAATATTTCCTGCCACGCGCTGGGTG 154482
 QY 4002 ctaaaacaagatgagatgagcgctaccactgaagctgaaagtgcgcattgaaacgga 4061
 Db 154481 CTRAAACAACAGATGAGGATGAGGTACCACTGAAGTCTGAAGATGTGCCTATGAACGGA 154422
 QY 4062 cagtgttttcatatgtttcttaggttctcttatgcttaacagtttccaaagccagcccccacag 4121
 Db 154421 CAGTGTTTTTCATATGTTTCTAGGTTGTCTTATGCTACAGTTTCCAAAGCCAGCCCCACAG 154362
 QY 4122 tsgagaaatgtgtgagggcacccgacacaaactgcaaatgtgtttttaaagtcgaagtgcac 4181
 Db 154361 TGAGGAATACTGTGAGGCACCGCACACAACTGCAATGTGTTTAAAGTCAAGGTGACAC 154302
 QY 4182 atgtatttaagattttttttaaatactcttttgcagttaaatctcacttttttcaaacaaag 4241
 Db 154301 ATGTATTAAAGATTTTATTTTAAATCTCTTTCAGTTAAATCTCACTTTTCAACAAG 154242
 QY 4242 cctggatcagggcacaaacacttatattgttttagctgagagctcagcagcgagattg 4301
 Db 154241 CCTGGATCAGGGCAAAACAACCTTATATTGTTTGTAGTCTGAGGCTCAGCAGGCGAGATTG 154182
 QY 4302 caggcaggggggcacttttcatcctagggggccagcctggtggcctggactctgacac 4361
 Db 154181 CAGCAGGGGGGCACCTTTTCATCCATGAGGGCCAGCCTGGGGCTGGGACTC-GATCAC 154123
 QY 4362 cattgtgagggccagggcagctgctgctatggaggagagaaatgcaaacctgaacgcaggttt 4421
 Db 154122 CATTTGTGGAGCCAGAGGACAGTGCCTGATGGAGGAGAAATGTCAAACGTGAACGCGAGTTT 154063
 QY 4422 caccactctaggaagcagcttattgagccctcagctgagatgtgttagagggatggg 4481
 Db 154062 CACCACCTCTAGGAAGCAGCTTGTGTAGCCCTGACCTGGATGTGTTAGAGGGATGGG 154003
 QY 4482 ctgaataggcaggttagatttctctgctcaacagctgcttgggaagctgtgtgattcct 4541
 Db 154002 CTGAATAGGCAGGTAGATTTCCTGCATCAACAGTGTCTTTGGGAAGCTGTGTGGATTCT 153943
 QY 4542 gaggaaagacagggagcggagatggagccacacatgagtttgcaccgctactgcagc 4601
 Db 153942 GAGGAAGAACAGGGAGGCGGAGATGGAGCCACACATGATGTTTGTCTACCCGCTACTGCGAGC 153883
 QY 4602 acttctgaccagaaatctcatgtcccaaaaccccatgtaaaccttcaacacacaaagct 4661
 Db 153882 ACTTTGTACCCAGAACTCAITGTCACAAACCCCATGTAAACTTCAACCCACTCAAGACT 153823
 QY 4662 gttatttcgctggaagaaataaaccttttttctcaccagctcatattgtaacctcttcata 4721
 Db 153822 GTTTATTTCGCTGAAGAAATAAATTTTTCACCCAGTCATTTTGTACCTCTCTCATAT 153763
 QY 4722 ggcgtgtgcacccctccagaaacgttggttatacttccagtcagtgctggagaaactgaag 4781
 Db 153762 GGCCTGTGCACCCCTCCAGAAACGTGGTTATATCTTCAGTCAAGTGTGGGAACTGAAG 153703
 QY 4782 acttccggttgcaggaactgaggttgaccttggggaagaaagttccacatcatctta 4841
 Db 153702 ACTTCCGGTTGGTCGAGGAACTGAGGTTGACCTTCGGGAAGGAAGTTCACATCATCTTA 153643
 QY 4842 ttattatgcctgtgattgtgggtcctgcagcagggagacatccagtaactcgggtgtcttcaat 4901
 Db 153642 TTTATTATGCTGTGATGTGGTCTGCCAGGAGACATCCAGTACTCGGTGCTTTAAT 153583
 QY 4902 tgcacctggggaactgtgtttatttgccotttcttggggcatacctggttttggatgaagt 4961

Db 153582 TGCCACCTGGGGAAGTGTGTTATTTGGCTCTTTGGGGCATCTGTGTTTGGATGAAGT 153523
 QY 4962 gaggggaatacagaggtaaaagaattgtctccaccctgaagcggggagtgcccgcttcaca 5021
 Db 153522 GAGGGGAATACAGAGTAAAGAAATTTCTCCACCCCTGAAGCGGGAGTCCCGCTTCACA 153463
 QY 5022 ttcttgaaatggtgagccactgggagacagttctgcccgggacatggttgcttctcaa 5081
 Db 153462 TTTCTGGAATGGTGCAGCCACTGGGACAGTCTGCGCCGGGCATGGTTGTTTTCAA 153403
 QY 5082 ggtccttaataataaccctattcttacataacccttgccctgagtggttttaagcaag 5141
 Db 153402 GGTCTCTAATATAATCCCTATTCTTACATAATCCCTTGCCCTTGATGGTTTAAAGCAAG 153343
 QY 5142 aactcctgtgtcccatggtctccaccactcaccatcaccctgctgtgagcaagagtcctag 5201
 Db 153342 AACTTCCCTGTGTCCCATGGTCTCCACCACCTCACCATCACCCTGCTGTAGCAAGAGTCTCTAG 153283
 QY 5202 tcaggggaggtgcaatttagtagttaattgcaacttatccatgagataaataaaagaga 5261
 Db 153282 TCAGGGAGGTGCATTTTAGTAGTTACATTGCACTTATCCATGAGATAAAATAAAGAGAGA 153223
 QY 5262 actgtttttatcagtgagggtcaacctaaaatttcaaaagtgcgcctttttgaaatcttg 5321
 Db 153222 ACTGTGTTTATCAGTGGAGGCTAACCTTAAATTTCAAGTGTGCGCTTTTGAATCTTG 153163
 QY 5322 ggcctctctctgtagaaacaaatggccctttgtggtcagcggcctgcacctaactggga 5381
 Db 153162 GGCCTCTCTCTGTAGAACCAATAGGCCCTTTGTGGTCTACGGCTCGCACCTAACCTGGA 153103
 QY 5382 gagtctgagctcctgagctcactgagccacagactaggtcttctgctcctccgc 5441
 Db 153102 GAGTTCTGAGCTCTCGAGCTCAGCTGAGCCACAGACTAGGCTTCTTGGCTCCTCCGC 153043
 QY 5442 agcatgctcctcaccgccagaaacccgcagctgtgggaagaccatgagggagctatt 5501
 Db 153042 AGCATGCTCTGCTCACCCTCAGAACCCGACGCTGTGGAGAGCCATCTAGGAGGCTATT 152983
 QY 5502 ccaaggatacacttcacactgccttcagctgagctcagctgacagctgacaaatactcctcta 5561
 Db 152982 CCAGGGATACATTCOCACCTGCCTTTCAGCTGAGCTCAGCTCAGCTGACAAATCTCTCTA 152923
 QY 5562 tcggagccagaagacttcagctccacaaatgaagtctctgctctgaaacacattcttgg 5621
 Db 152922 TCGGAGCCAGAGACTTCAGCTCCACAAATGAAGTGTCTGTCTCCTGAAACATCTCTGG 152863
 QY 5622 gaagaatcccaacatcgagaaaacggtgctcctgtgagttcccaacaatgcttcttctcat 5681
 Db 152862 GAAGAAATCCCAACATCAGAAAAACGGTCTCTGTGAGTTCCAAACATGCTTCTTGTTCAT 152803
 QY 5682 ggttctctcgtatgagtggaattgaagatggtttatttattgttcttaactgagaaaa 5741
 Db 152802 GGGTCTCTCCGATGAGGTGGATTAAGAGTGTGTTTATTTGTTGTTCTAACTGAGAAAA 152743
 QY 5742 aaaggaggcccccacaaaggttgaggtcacacagctctccacagtttccagagggcggttgg 5801
 Db 152742 AAAGGAGCCACCCACAAAGTTGAGGTACACAGTCTCCACAGTTCCAGAGGGGTTTGG 152683
 QY 5802 ggttggggaaggcacctccagagatgaggtccttaaggggacatgagtaagcatgctg 5861
 Db 152682 GGGTGGGGAAGGACCTCCAGAGCATGAGGCTCTAAGGGGACATGAGTAAAGCATGCTG 152623
 QY 5862 tgaccagtggaaggagagccagctcactcctcgacggggttcctagctgagaaa 5921
 Db 152622 TGACCCAGTGAGGAAGGAGGAGCCAGCTGCATCTCTGACGGGGTTCTTAGCTGCAAGAA 152563
 QY 5922 ggttcccgcttagccgaggggaaacaccttgatagcaagaagagccctggatgacacctg 5981
 Db 152562 GGGTCCCGCTTAGCCGAGGGGAAACACCTTGATAGCAGAGAGGCGCTGGATGCACACTG 152503
 QY 5982 gacgcgaggtctcgcgccagacacagtgctccatgtcagccctgcacctggggtgt 6041

Db 152502 GCACGCCGAGGCTCTCGCCCCAGACACAGTGTCTCATGTCTACGCCCTGCACCTGGGGTGT 152443
 QY 6042 gtgattcactgacagatgccacaatctctgcaccaatataccacagatgggggaaagtgt 6101
 Db 152442 GTGATTTCAGCTGCACAGATGCCAATCTCTGCACCAATATCCACAGATGGGGAAGGTG 152383
 QY 6102 agaggaaggggcaagtgtgttaactgctcaagagatgcttaaacctccataagagaga 6161
 Db 152382 AGAGGAAGGGGCAAGTGTGTAACTGCTCAAGAGATGCTTAAACCTCCATAGAGAGGA 152323
 QY 6162 gcccgggcaggggacatctgtgtcccgctcacacactgcagcaggggaaggtggtctggc 6221
 Db 152322 GCCGGGCGCAGGGGCATCTGTGTCTCCGTTCACACTGCAGCAGGGAAGGTGGCTGGC 152263
 QY 6222 tggctcctggcactcagtggtgttgaagctccagaggggtcttattgacctgtcttt 6281
 Db 152262 TGGCTCCCTGGCATCAGTGGTTTGGTTTAAAGTCCAGAGGGTCTTATTGCCATTGTCTTT 152203
 QY 6282 tcctctgccccctgagccagcctaaggccctggagctgtctgtttctttaggcggatgaaactg 6341
 Db 152202 TCCTCTGCCCTTTGAGCCAGGCTTAAGGCCCTGGAGTCTGTGTTCTTTAGGCCGATGAACGTG 152143
 QY 6342 acatgctcctaccatgaccaggtctgtggcaaggctcctcactcagctatctcttgaggtgg 6401
 Db 152142 ACATGCTCTACCATGACCAGGCTCTGGGCAAGGCTCTCTCACAGTATCTTTGAGAGGTGG 152083
 QY 6402 gcatggagtgccccattctcaggtacagaaaccttcagagaggaataaataagcttgccct 6461
 Db 152082 GCATGGAAGTGCCCATTTCTCAGGTACAGAAACCTTCAGAGAGGATAAATAGCTTGCCCT 152023
 QY 6462 gtagaacagagactgaacacctgtccgctgactcctccagctactctgcccactgtag 6521
 Db 152022 GTAGAACAGAGACTGAACCCCTTGTCTGCCGTGACTCCCCAGCTACTCTGCCACTGTAG 151963
 QY 6522 cccctgcttactgtcctgtgacacccctcaccatcctgtatacctttaaataataatttt 6581
 Db 151962 CCCCTGCCTTACTGTCTTGGCACACCCCTCACCATCTCTGTATACCTTAAATATCAAAGA 151903
 QY 6582 gggcaagagagaaaggcctttaagataaagtatttttttaaggaaccttaattatttt 6641
 Db 151902 GGGCAAGAGAGAAAGGCTTTAAAGATAAGTTATTTTTTAAAGAACCTTAAATATTATTT 151843
 QY 6642 ttaagaagtacacaaatagtgactgaaatgc-aaaaaataaaaaaaatgctgact 6700
 Db 151842 TTAGAAGTAAACAAATTAGTACGCTGAATGCAAAATAAAAAAATAAATGCTGACT 151783
 QY 6701 accctttgaaaatgcttctcagattgtttttatatgtaatctctagacacttgctcat 6760
 Db 151782 ACCCTTTGAAAATGTGCTTTCAGATTGTTTTTATATGTAAATCTTAGACACTTGTCTAT 151723
 QY 6761 taagaaaatagtggtgctgtgctcagcaagaagcacactggcagctggtggttggtat 6820
 Db 151722 TAAGAAAATAGTGTGTGTGTCTCAGCAAGAACACACTGGCAGCTGGCTTTGGTAT 151663
 QY 6821 aggaagtggaggcaagacctgggtttctgacaagtgcctcagacttacccttccatc 6880
 Db 151662 AGGAAGTGAAGGCAAGGACCTGGGTTCTTGACAAGTGCAGTCAAGATACACATCCATC 151603
 QY 6881 tgagagctggtgcttgtgctccctgggttagggc 6915
 Db 151602 T-GAGAGCTGCTTGCCTTGTTCCTCGTGTGAAGGCC 151569

RESULT 4
 AC092983_0
 WPCOMMENT
 Sequence split into 4 fragments LOCUS AC092983 Accession AC092983
 Fragment Name Begin End
 AC092983_0 1 110000
 AC092983_1 100001 210000
 AC092983_2 310001 350000
 AC092983_3 350001 350921
 LOCUS AC092983 350921 bp DNA linear HTG 12-JAN-2002

Qy	4782	acttccggttggtcgaggaaactgagggttgaccttcgggaaggagttccactcatctta	4841
Db	78479		
		acttccgggttggttcaggaaactgagggtgaccttccgggaaggagttccactcatctta	78538
Qy	4842	ttattatgcctctgtagtgggtcctcgaggggagacatccagtactcgtggtctttaat	4901
Db	78539		
		tttatattgctctgtagtgggtcctcgaggggagacatccagtactcgtggtctttaat	78598
Qy	4902	tgcacctggggaactgtgtttattgacctctcttggggcaacctggttttgggatgaagt	4961
Db	78599		
		tgcacctggggaactgtgtttattgacctctcttggggcaacctggttttgggatgaagt	78658
Qy	4962	gaggggaatacagagggtaaaaaattgtctccacctgaagcggggagtcctccaca	5021
Db	78659		
		gaggggaatacagagggtaaaaaattgtctccacctgaagcggggagtcctccaca	78718
Qy	5022	ttcttgaaaatggtgcagccaactgggacagttctctcccgggcatagttgtttctcaa	5081
Db	78719		
		tttcttgaaaatggtgcagccaactgggacagttctctcccgggcatagttgtttctcaa	78778
Qy	5082	ggctcttaataataatocctattcttacaataaccttggccctgtagtgttttaagcaag	5141
Db	78779		
		ggctctctaaataataatocctattcttacaataaccttggccctgtagtgttttaagcaag	78838
Qy	5142	aactcctgtgtcccataggtctccaccactcaaccataacctgtctgtagaagagtcctag	5201
Db	78839		
		aactcctgtgtcccataggtctccaccactcaaccataacctgtctgtagaagagtcctag	78898
Qy	5202	tcaggggagggtgcatttttagtagttaattgcacttatccatgagataaaataaaagaga	5261
Db	78899		
		tcaggggagggtgcatttttagtagttaattgcacttatccatgagataaaataaaagaga	78958
Qy	5262	actgtttttatcagtgagggtcaacctaaaaattcaaaagtgtcgcccttttgaactcttg	5321
Db	78959		
		actgtttttatcagtgagggtcaacctaaaaattcaaaagtgtcgcccttttgaactcttg	79017
Qy	5322	ggcctctctctctagaaccaaaggccctttgtgtctcaggcctcgcaactaacctgga	5381
Db	79018		
		ggcctctctctctagaaccaaaggccctttgtgtctcaggcctcgcaactaacctgga	79077
Qy	5382	gagttctgagctctcgagctcaacctgagccacagactagggtcttcttgctcctccgc	5441
Db	79078		
		gagttctgagctctcgagctcaacctgagccacagactagggtcttcttgctcctccgc	79137
Qy	5442	agatgctgtctacccccagaaaccccgagctgtgggaagagccatgtaggaggctatt	5501
Db	79138		
		agatgctgtctacccccagaaaccccgagctgtgggaagagccatgtaggaggctatt	79197
Qy	5502	cccaggcatcacattccactgccttcagctgagctcacagctgcaaatcatctcccta	5561
Db	79198		
		cccaggcatcacattccactgccttcagctgagctcacagctgcaaatcatctcccta	79257
Qy	5562	tcggagccagaagacttcagctccacaaaaatgaagtctctgtcctgaaaaactcttgg	5621
Db	79258		
		tcggagccagaagacttcagctccacaaaaatgaagtctctgtcctgaaaaactcttgg	79317
Qy	5622	gaagaaatcccaacatcgagaaaaacggtgtcctgtgagttccacaacatgtctcttcat	5681
Db	79318		
		gaagaaatcccaacatcgagaaaaacggtgtcctgtgagttccacaacatgtctcttcat	79377
Qy	5682	gggtttcttcogtatgagtggaataagagtgttttatttctgttctcaactgagaaaa	5741
Db	79378		
		gggtttcttcogtatgagtggaataagagtgttttatttctgttctcaactgagaaaa	79437
Qy	5742	aaaggaggccaccacaagggttgaggtcacacagttctccacagtttccaggagcggttgg	5801
Db	79438		
		aaaggaggccaccacaagggttgaggtcacacagttctccacagtttccaggagcggttgg	79497
Qy	5802	gggttggggaaggcaacctccagagacatgaggtcttaaggggacatgagtaaaagatgtctg	5861
Db	79498		
		gggttggggaaggcaacctccagagacatgaggtcttaaggggacatgagtaaaagatgtctg	79557

Qy	5862	tgaccacagtaggaagaagacgcagctgcaactctgcacgggttccctagctcagaa	5921
Db	79558	TGACCCCAAGTAGGAAGGAGAGGCCAGCTGCACCTCTGCACGGGTCTCTAGCTGCAGAA	79617
Qy	5922	gggtcccgcttagccqaggggaaacacctgtagcagaagagcgctggatgcacacctg	5981
Db	79618	GGGTCCCGCCTAGGCCGAGGGGAAACACCTGATAGCAAGAAGGCCCTGGATGCACACCTG	79677
Qy	5982	gcacgcgaggtctccgccacagacacacagtgctccatgtcagccccctgcacctggggtgt	6041
Db	79678	GCACGCGAGGCTCTCGCCCGACACACAGTGTCTCCATGTACAGCCCTGCACCTGGGGTGT	79737
Qy	6042	gtgattcacgtgcacagatgcacaatcctgcaccaaatccacacagatgggggaagtgt	6101
Db	79738	GTGATTCACTGTGCACAGATGCCAAATCTTGCAACCAATATCCACAGATGGGGGAAGGTG	79797
Qy	6102	agagaaagggcaagtgtatgttaaactgctcaagatgcttlaaacctccatagagagga	6161
Db	79798	AGAGAAAGGGCAAGTGTATGTAACTGCTCAAGAGATGCTTAAACCTCCATAGAGAGGA	79857
Qy	6162	gccggggcgagggcatctgtgtgtcccgctcacacaactgcagcaggaaggcgctggc	6221
Db	79858	GCCGGGGCGAGGGCATCTGTGTGCCGTCAACACTCTACGACAGGAAGGTGGCTGGC	79917
Qy	6222	tggctccctggcatcagtggttgggtttaagctccagagggctctattggcaattgcttt	6281
Db	79918	TGGCTCCCTGGCATCAGTGGTGTGGTTTAAGCTCCAGAGGGCTTATTATGGCCATTGCTCTT	79977
Qy	6282	tcctctgcccccttgagccagcctaaaggccccggagctgtcttcttttagcgagatgaactg	6341
Db	79978	TCCCTCGCCCTTGAGCCAGCCCTAAGGCCCTGGAGTCTCTTTTAGCGGATGAACCTG	80037
Qy	6342	acatgctctaccatgacagagctctggcaaggctcctcacagtatccttgagagggtg	6401
Db	80038	ACATGCTCTTACCATGACAGGCTCTGGCAAGGCTCTCACAGATATCCTTGAGAGGTGG	80097
Qy	6402	gcattggaagtggccattcttcaggtacagaaacccctcagagggagataaatagcttgccct	6461
Db	80098	GCATGGAAGTGCCATTCTCTCAGTACAGAAACCTTCAGAGAGGATAAATAGCTTGCCCT	80157
Qy	6462	gtagaacagagactgaacacctgttcgcgctgactcccccaagctactgtcccactgtag	6521
Db	80158	GTAGAAGCAGGACTGAACCCCTTGTCCGGCTGACTCCCCCAGCTACTTGCCCCACTGTAG	80217
Qy	6522	ccccctgcttactgtctgtgcacacccctcaccatctctgtatccttaaatatcaaaaga	6581
Db	80218	CCCCCTGCCTTACTGTCTTGGCACACCCCTCACCATCCTGTATACCTTAAATATCAAGA	80277
Qy	6582	ggccaagagagaaaggctttaagataagttattttttaagggaaccttaatatattt	6641
Db	80278	GGGCAAGAGAGAAAGGCTTTAAAGATAAGTTATTTTTTTTAAGGAACCTTAAATATATT	80337
Qy	6642	ttaagaagtaaccaaatagtgtacgtgaaatgc-aaaaaaaaaaaaaaaaaagtctgact	6700
Db	80338	TTANGAAGTAACCAATATTAGTACGTGAATTCGCAAAAAAAAAAAAAAAAAAATGCTGACT	80397
Qy	6701	accttttgaattgtgtcttcagatgttttttatatgttaattcttagacacttgtcat	6760
Db	80398	ACCCCTTTGAAAAATGTCTTTCAGATGTGTTTTATATGTAATCTTAGACACTTGTCTAT	80457
Qy	6761	taagaaatagttgcttggtctgtgtctgcagaagacacactggcacgtggtcttggtat	6820
Db	80458	TAGAAAAATAGTGGCTGGCTTGTGCTCAGCAAGAAGCACACTGGCACGCTGGCTTTGGTAT	80517
Qy	6821	aggaagtggaaagcaaggaacctgggtttctgcagaagtgcgttcagacttacccttccatc	6880
Db	80518	AGGAAGTGGAAAGCAAGGACCTGGGTCTGTGACAAGTGCAGTCACATACACATCCATC	80577
Qy	6881	tggagaactgggtgtgcttggctccccctggtagggc	6915
Db	80578	T-GAGAGCTGCTTGGCTTGTGTTCCTGGTGAAGGCC	80611


```

Db 789 ACAGTGGGGAAGGTGAGAGGAGGCGCAAGTGATGTAACTGCTCAAGAGATGCTTA 848
QY 6145 aacctccatagaagagcagggcagggcagcctctgtgtcccgccacacactgcagc 6204
Db 849 AACCTCCATAGAGAGGAGCGGGCGAGGGCATCTGTGTGTCCTCCGTCACACATCGCAGC 908
QY 6205 agggaagggtggctggctgctccctggcctcagtggtttggttttaagctccagagggtc 6264
Db 909 AGGGAAGGGTGGCTGGCTGCTCCCTGGCATCAGTGGTGGTTTAAGCTCCAGAGGTTTC 968
QY 6265 ttatgcatgttcttctctctgccccctgagccagcctaaagggccctggagtgctgttc 6324
Db 969 TTATTGTCATTGTCTTTTCTCTGCTCCCTTGGAGCCAGCCCTAAGGCGCCTGGAGTCTGTTTC 1028
QY 6325 tttagcgagatgaactacatgctctaccatgacacagggcctctgggcaagcctccctca 6384
Db 1029 TTTAGGCGGATGAACATGACATGCTCTTACCATGACAGGCTCTGGGCAAGGCTCCTCACA 1088
QY 6385 gtatccttgagagtgggcagtggaagtgcacattctcaggtacagaaaccttcagagag 6444
Db 1089 GTATCCTTGAGAGTGGGCGATGGAAGTGCCATTCTCAGGTACAGAAACCTTCAGAGAG 1148
QY 6445 gataaatagcttgccctgtagaagcagtgactgaacaccttgcgcctgactccccccagc 6504
Db 1149 GATAAATAGCTTGCCCTGTAGAAGCAGGACTGAAACCCCTTGTTCGGCCTGACTCCGCCAGC 1208
QY 6505 tactctgccactgtaagccctcttactgtctctgagcaccctccaccatctcttat 6564
Db 1209 TACTTCTCCCACTGTAGCCCTGCTGCTTACTGTCTGGCACACCCCTCACCATCTCTAT 1268
QY 6565 accttaatatcaagaggcgcaagagagaaaggcctttaaagataaagtatttttttaag 6624
Db 1269 ACCTTAATATCAAGAGGCGCAAGAGAGAAAGGCTTTAAAGATAAGTATTTTAAAG 1328
QY 6625 gaaccttaattatttttaagaagtaaccacaaattagtcagtgaaatgcacacacacac 6684
Db 1329 GAACCTTAATATATTATTTTAAAGAGTAACCAAAATTAGTACGCTGAATGCAAAAAA 1388
QY 6685 aaaaaaa 6692
Db 1389 AAAAAAA 1396

RESULT 7
AC092983_0/c
WPCOMMENT
Sequence split into 4 fragments LOCUS AC092983 Accession AC092983
Fragment Name Begin End
AC092983_0 1 110000
AC092983_1 100001 210000
AC092983_2 200001 310000
AC092983_3 300001 350921
LOCUS AC092983 350921 bp DNA linear HTG 12-JAN-2002
DEFINITION Homo sapiens chromosome 3q clone RP11-717C3, WORKING DRAFT
SEQUENCE, 49 unordered pieces.
ACCESSION AC092983.2 GI:18139248
VERSION AC092983
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 350921)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Anaratinge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,

```

```

Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Prantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gili,R., Gorrell,J.H., Guevara,W., Gunaracne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulseghe,H., Lozado,R., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Noyan,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogulu,H., Okuwonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
Sisson,I., Sodergren,E., Sonaie,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczek,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE
JOURNAL
REFERENCE
2 (bases 1 to 350921)
AUTHORS
Worley,K.C.
TITLE
JOURNAL
Direct Submission
Unpublished
COMMENT
Submitted (09-AUG-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jan 12, 2002 this sequence version replaced gi:15136886.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HDEZ
Center clone name: RP11-717C3
----- Summary Statistics
Assembly program: Phrap; version 0.990329first call to
findPhrapList
Consensus quality: 333925 bases at least Q40
Consensus quality: 347288 bases at least Q30
Consensus quality: 355437 bases at least Q20
Estimated insert size: 353663; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 3.6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 49 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 50319: contig of 50319 bp in length
* 50320 50419: gap of unknown length
* 50420 72332: contig of 21913 bp in length
* 72333 72432: gap of unknown length
* 72433 94554: contig of 22122 bp in length
* 94555 94654: gap of unknown length

```

* 94655 117005: contig of 22351 bp in length
* 117006 117105: gap of unknown length
* 117106 133005: contig of 16200 bp in length
* 133006 133405: gap of unknown length
* 133406 145261: contig of 11856 bp in length
* 145262 145361: gap of unknown length
* 145362 155243: contig of 9882 bp in length
* 155244 155343: gap of unknown length
* 155344 163670: contig of 8327 bp in length
* 163671 163770: gap of unknown length
* 163771 171407: contig of 7637 bp in length
* 171408 171507: gap of unknown length
* 171508 179889: contig of 8382 bp in length
* 179890 179989: gap of unknown length
* 179990 187364: contig of 7375 bp in length
* 187365 187464: gap of unknown length
* 187465 194762: contig of 7298 bp in length
* 194763 194862: gap of unknown length
* 194863 203434: contig of 8572 bp in length
* 203435 210606: contig of 7072 bp in length
* 210607 210706: gap of unknown length
* 210707 216745: contig of 6039 bp in length
* 216746 216845: gap of unknown length
* 216846 222655: contig of 5810 bp in length
* 222656 222755: gap of unknown length
* 222756 228149: contig of 5394 bp in length
* 228150 228249: gap of unknown length
* 228250 233372: contig of 5123 bp in length
* 233373 233472: gap of unknown length
* 233473 238412: contig of 4940 bp in length
* 238413 238512: gap of unknown length
* 238513 244488: contig of 5976 bp in length
* 244489 244588: gap of unknown length
* 244589 250356: contig of 5768 bp in length
* 250357 250456: gap of unknown length
* 250457 255459: contig of 5003 bp in length
* 255460 255559: gap of unknown length
* 255560 260074: contig of 4515 bp in length
* 260075 260174: gap of unknown length
* 260175 266047: contig of 5873 bp in length
* 266048 266147: gap of unknown length
* 266148 270001: contig of 3854 bp in length
* 270002 270101: gap of unknown length
* 270102 275022: contig of 4921 bp in length
* 275023 275122: gap of unknown length
* 275123 278889: contig of 3767 bp in length
* 278890 278989: gap of unknown length
* 278990 282566: contig of 3577 bp in length
* 282567 282666: gap of unknown length
* 282667 286119: contig of 3453 bp in length
* 286120 286219: gap of unknown length
* 286220 289575: contig of 3356 bp in length
* 289576 289675: gap of unknown length
* 289676 294061: contig of 4386 bp in length
* 294062 294161: gap of unknown length
* 294162 298031: contig of 3870 bp in length
* 298032 298131: gap of unknown length
* 298132 303266: contig of 5135 bp in length
* 303267 303366: gap of unknown length
* 303367 305907: contig of 2541 bp in length
* 305908 306007: gap of unknown length
* 306009 308221: contig of 2214 bp in length
* 308222 308321: gap of unknown length
* 308322 310563: contig of 2242 bp in length
* 310564 310663: gap of unknown length
* 310664 313552: contig of 2889 bp in length
* 313553 313652: gap of unknown length
* 313653 316728: contig of 3076 bp in length
* 316729 316828: gap of unknown length
* 316829 319666: contig of 2838 bp in length
* 319667 319766: gap of unknown length
* 319767 323675: contig of 3909 bp in length

* 323676 323775: gap of unknown length
* 323776 327559: contig of 3784 bp in length
* 327560 327659: gap of unknown length
* 327660 331218: contig of 3559 bp in length
* 331219 331318: gap of unknown length
* 331319 333580: contig of 2262 bp in length
* 333581 333680: gap of unknown length
* 333681 336280: contig of 2600 bp in length
* 336281 336380: gap of unknown length
* 336381 338607: contig of 2227 bp in length
* 338608 338707: gap of unknown length
* 338708 341633: contig of 2926 bp in length
* 341634 341733: gap of unknown length
* 341734 344669: contig of 2936 bp in length
* 344670 344769: gap of unknown length
* 344770 348238: contig of 3469 bp in length
* 348239 348338: gap of unknown length
* 348339 350921: contig of 2583 bp in length.
FEATURES
 Location/Qualifiers
 1. .350921
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="3q"
 /clone="RP11-717C3"
BASE COUNT 97816 a 75489 c 76879 g 95821 t 4916 others
ORIGIN
Query Match 19.6%; Score 1359.8; DB 2; Length 110000;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1367; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
Qy 936 ttcaatcgctgggattagctacggtcaagtgcgtggcacagctattgaacaagagacttc 995
 |||||
Db 41663 TACAGTCGCTGGGATTAGCTACGCTCAAGTCGTCGTCACACACTATTGAACAAGACTTC 41604
Qy 996 cagcgaccacacagaccacacctaccctgtcatctacttccaccaaaaggagacgggctt 1055
 |||||
Db 41603 CAGCGACACACAGACACACACTACCTGTCTACTTCTTCCACCAAGAGAGACGGCGTT 41544
Qy 1056 actgtccattacagataaacagttcctcctcagacatgtgagagctcaactcttatat 1115
 |||||
Db 41543 ACTGTCCATTACAGATAACAGTTCTCTCTCAGACATTTGTGGAGAGCTCAACTTCTTATAT 41484
Qy 1116 taaatctcaaacctcttcacattcagagatattctctctctctcagctcagagtgagag 1175
 |||||
Db 41483 TAAATCTCAAACCTCTTCACATTCAGAGTATTCCTCTCTTTTCATGCTCAGACTGAGAG 41424
Qy 1176 aagtaacatctcctcctatgacgggggaatatgctcagcctctcctactgagtcgacgttct 1235
 |||||
Db 41423 AAGTAACATCTCATCTTATGACGGGGGAATATGCTCAGCCTTCTTACTGAGTCGCCAGTTCT 41364
Qy 1236 gcatacatcaaaccttcgctcctacacacacccattaaatgcgaacaccttcggttgt 1295
 |||||
Db 41363 GCATACATCAAACTTCCTCGCTTACACACCCACCATTAAATATGCGAACACTTCGGTGT 41304
Qy 1296 tctggacactgactgagttgttagtgactcctcctcctcctcctcctcctcctcctcctc 1355
 |||||
Db 41303 TCTGGACACTGATGCTGAGTTGTTAGTGATCCCTCTCTCTCTCTCTCTCTCTCTCTCTC 41244
Qy 1356 ttcttctctcaggggctcctcttgcctctgcctctgctcctctgctcctcctcctcctcctcct 1415
 |||||
Db 41243 TTCTTCTTCTTACGGGCTCTCTTTCCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 41184
Qy 1416 ttcatcaattttaccatcaaacaggcctctgtgcatctactaactaagctcactcctgatgc 1475
 |||||
Db 41183 TTTCATCAATTTTACCATCAACAGGGCTCTGTGCATCTTACTAAGTCTTACTCTGTATGC 41124
Qy 1476 atccacaccatggtcttcctcaccatcacctttaccagatcccttaacgacatcatc 1535
 |||||
Db 41123 ATCCACACCATGGTCTTCTCTCACCATCACCTTTTACCAGTATCTCTTACGACATCATC 41064
Qy 1536 tggccactttctgtctcacaacaaccttgccacagtcactctctcctcctcctcctcctcctcct 1595
 |||||

Db 41063 TGCCCCACTTTTGTCTCACAACAAACCTTGCCACAGTCATCTTCTACCCCTGTCTCTGCC 41004
Qy 1596 cagggaagagagactctgtgacttcatttcagagacatcaacaatgacatcattcatgac 1655
Db 41003 CAGGCAAGGAGAGACTCTCTGTGACTTCTATTTTCAGACATCAACAATGACATCATTCATGAC 40944
Qy 1656 aatgtccatagtagtcaaaactgcagaccttaagagccagagcagcccccacaccacagagaa 1715
Db 40943 AATGCTCCATAGTAGTCAAACTGCAGACCTTAAGAGCCAGAGACCCCCACACCAAGAGAA 40884
Qy 1716 agtcattacagaatcaaaagtccaaagcctggtgtctctgtccacagagtcacccaagc 1775
Db 40883 AGTCATTACAGAAATCAAAAGTCACCAAGCCTGGTGTCTCTGCCACAGAGTCCACCAAGC 40824
Qy 1776 tgtacaacaactctcttgcctccatcttaacagagtcctccacagagcaaacct 1835
Db 40823 TGTAAACAACAATCTCTCTTGGCTCCATCTTAAACAGAGTCTCCACAGAGCAAAACCT 40764
Qy 1836 tccagccaaagcaccacttagcacaatgtctccaaacttcacaactaccattctgaa 1895
Db 40763 TCCAGCCACAGCACCACCTTAGCACAATGTCTCCAATTTCACAACTACCATCTCTGAA 40704
Qy 1896 gacctctcagcctcttatgacctctctgacctgtcaagcagcagcatctctggtcac 1955
Db 40703 GACCTCTCAGCCTCTTATGACCACCTCTCTGCGACCCCTGTCAAGCAGCAGCATCTCTGTCAC 40644
Qy 1956 tggccctatagcctacagactacagcttggaacagctctcgtgacccatcctgaaat 2015
Db 40643 TGGCCCTATAGCCGTACAGACTACAGCTGGAACAGCTCTCGTGCACCCATCTCTGAAAT 40584
Qy 2016 actagttctcaaatctcaacagaagtggtgcatcagcagacaagaagaccgagtgattgt 2075
Db 40583 ACTAGTTCTCAATCTCAACAGAAGTGGCATCAGCACAAGAAAGAACCCAGTGATTGT 40524
Qy 2076 ggatgtaccactggattgattccctttgaccagtgatcccatcagcaaaagaatgac 2135
Db 40523 GGATGCTACCACTGGATTGATCCCTTTGACCAGTGTACCCACATCAGCAAAAGAAATGAC 40464
Qy 2136 cacaagcttgcttacagcagagtaacagccagcttcacgttccctcaggaacatctcc 2195
Db 40463 CACAAAGCTTGGCGTTACAGCAGAGTACAGCCACAGCTTCAGCTTCCCTTCGGAACATCTCC 40404
Qy 2196 ttctcccaaacaccagctgtttccagcgtgagactgtgctcccaaatctgccacctt 2255
Db 40403 TTCTCCCAAACACCACAGTTGTTTCCACGGCTGAAGACTTGGCTCCCAATCTGCCACCTT 40344
Qy 2256 tgcgttcagagcagcacacagtcacacacacactgctccttcttcagcctcagtcacaa 2314
Db 40343 TGCTGTTTCAGAGCAGCACACAGTACCAACACAGTGTCTCTTCAGCCTCAGGTAATAA 40285

RESULT 8
AC026629 AC026629 118407 bp DNA linear HTG 04-JUN-2000
LOCUS Homo sapiens chromosome 3 clone RP11-816I3 map 3, LOW-PASS SEQUENCE
DEFINITION
SAMPLING.
ACCESSION AC026629
VERSION AC026629.2 GI:8247896
KEYWORDS HTG; HTGS-PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 118407)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, F.,
Boguslavsky, L., Bouckhagter, B., Brown, A., Burkett, G.,
Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D.,
Gallagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Hearford, A., Horton, L.,
Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J.,
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,
Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J.,
Vassiliou, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 4, 2000 this sequence version replaced gi:7284654.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: W1BR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8653
Center clone name: 816_I_3

* NOTE: this record contains 141 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 757: contig of 757 bp in length
* 758 857: gap of 100 bp
* 838 1584: contig of 727 bp in length
* 1585 1684: gap of 100 bp
* 1685 2449: contig of 765 bp in length
* 2450 2549: gap of 100 bp
* 2550 3306: contig of 757 bp in length
* 3307 3406: gap of 100 bp
* 3407 4163: contig of 757 bp in length
* 4164 4263: gap of 100 bp
* 4264 5024: contig of 761 bp in length
* 5025 5124: gap of 100 bp
* 5125 5878: contig of 754 bp in length
* 5879 5978: gap of 100 bp
* 5979 6734: contig of 756 bp in length
* 6735 6834: gap of 100 bp
* 6835 7583: contig of 749 bp in length
* 7584 7683: gap of 100 bp
* 7684 8446: contig of 763 bp in length
* 8447 8546: gap of 100 bp
* 8547 9304: contig of 758 bp in length
* 9305 9404: gap of 100 bp
* 9405 10161: contig of 757 bp in length
* 10162 10261: gap of 100 bp
* 10262 11003: contig of 742 bp in length
* 11004 11103: gap of 100 bp
* 11104 11831: contig of 728 bp in length
* 11832 11931: gap of 100 bp
* 11932 12694: contig of 763 bp in length
* 12695 12794: gap of 100 bp

Db 84482 ACAATGCTTCTGTTTCATGGGTTTCTTCCTGATGGAGTGAAGAGTGTATTTATTTG 84541
QY 5724 ttgttctaactgagaaaaagaggagcaccacaaggttgaggtcacacagctccacag 5783
Db 84542 TTGTTCTTAAGTGAAGAAAAAGAGGAGGACCCACAAAGGTTGAGGTTCACACAGTCTCCACAG 84601
QY 5784 ttccagagagcgtttggggtggggaagcactccagagcatgagctcgaaggagac 5843
Db 84602 TTTCCAGAGCGTTTGGGGTGGGGAAGGACCTCCAGAGCATGAGGCTTAAGGGGAC 84661
QY 5844 atgagtaaacatgtctgtgaccacagtgaggaagggagagggcagctgcactcctgcacy 5903
Db 84662 ATGAGTAAGCATGCTGTGTGACCCANTGAGGAAGGGAGAGGCCAGCTGCACCTCCTGCACG 84721
QY 5904 gggttctaactgagaaaggtcccgctaggccaggaagggaaacacatgtagcagaaga 5963
Db 84722 GGGTTCCTAGCTGAGAAAGGTCCTCGCTTANGCGGAGGGGAAACACCTGATAGCAGAAAG 84781
QY 5964 ggctggtatgcacacctggcagcgcaggtctctccgcccagacacagctgctccatgctcag 6023
Db 84782 AGCTGGATGCACACCTGGCAGCGGAGGCTCTNGCGCCAGACACAGCTGTNNNNNNN 84841
QY 6024 ccctgcacctggggtgtgtgattacagtcgacagatgycacaatcctgcaccaatctcc 6083
Db 84842 NNN 84901
QY 6084 cacagatgggggaaggtgagagaaggggcaag 6116
Db 84902 NNN 84934

RESULT 9
AC026629/c
LOCUS
DEFINITION Homo sapiens chromosome 3 clone RP11-81613 map 3, LOW-PASS SEQUENCE
ACCESSION AC026629
VERSION AC026629.2 GI:8247896
KEYWORDS HTG: HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 118407)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 3, clone RP11-81613
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 118407)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bida,F.,
Bohuslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodgson,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galligan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hages,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRoque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
TITLE
JOURNAL
Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Jun 4, 2000 this sequence version replaced gi:7284654.

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8653
Center clone name: 816_I_3

* NOTE: This record contains 141 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 757: contig of 757 bp in length
* 758 857: gap of 100 bp
* 858 1584: contig of 727 bp in length
* 1585 1684: gap of 100 bp
* 1685 2449: contig of 765 bp in length
* 2450 2549: gap of 100 bp
* 2550 3306: contig of 757 bp in length
* 3307 3406: gap of 100 bp
* 3407 4163: contig of 757 bp in length
* 4164 4263: gap of 100 bp
* 4264 5024: contig of 761 bp in length
* 5025 5124: gap of 100 bp
* 5125 5878: contig of 754 bp in length
* 5879 5978: gap of 100 bp
* 5979 6734: contig of 756 bp in length
* 6735 6834: gap of 100 bp
* 6835 7583: contig of 749 bp in length
* 7584 7683: gap of 100 bp
* 7684 8446: contig of 763 bp in length
* 8447 8546: gap of 100 bp
* 8547 9304: contig of 758 bp in length
* 9305 9404: gap of 100 bp
* 9405 10161: contig of 757 bp in length
* 10162 10261: gap of 100 bp
* 10262 11003: contig of 742 bp in length
* 11004 11103: gap of 100 bp
* 11104 11831: contig of 728 bp in length
* 11832 11931: gap of 100 bp
* 11932 12694: contig of 763 bp in length
* 12695 12794: gap of 100 bp
* 12795 13547: contig of 753 bp in length
* 13548 13647: gap of 100 bp
* 13648 14410: contig of 763 bp in length
* 14411 14510: gap of 100 bp
* 14511 15268: contig of 758 bp in length
* 15269 15368: gap of 100 bp
* 15369 16119: contig of 751 bp in length
* 16120 16219: gap of 100 bp
* 16220 16986: contig of 767 bp in length
* 16987 17086: gap of 100 bp
* 17087 17816: contig of 730 bp in length
* 17817 17916: gap of 100 bp
* 17917 18663: contig of 747 bp in length
* 18664 18763: gap of 100 bp
* 18764 19505: contig of 742 bp in length
* 19506 19605: gap of 100 bp
* 19606 20362: contig of 757 bp in length
* 20363 20462: gap of 100 bp
* 20463 21180: contig of 718 bp in length
* 21181 21280: gap of 100 bp

Db	31390	CCTCGGAGAACATCAGCTTGCCAGCAGCTCTGAGGTGCAAAATGGAATCCCATGTCTCA	31449
Qy	666	gactgacactgtctgactaggtcagtcgcaccatgaggtggagagatactgacacactg	725
Db	31450	GACTTGACAGCTGTGCTAGGTGGTTCGCCACCATCAGAGGTGGAGAGATCACTGTACACTG	31509
Qy	726	gctcttgaccacagcacaacatctgcagatgtgacaggaagctctgtcttcatactctga	785
Db	31510	GCTCTTGACCAACAGCACACATCTGCAGATGTGCACAGGAAGCTCTGCTTCATATCCTGA	31569
Qy	786	aggtgtgaatgcttcagttgttgacccagttctcagactctactgtacagttctggagaaag	845
Db	31570	AGGTGTGAATGCTTCACTGTGTACCCAGTGTCTCAGACTCTACTGTACAGCTGTGGAGAAAG	31629
Qy	846	tcacacagcatggagatagaggatttaccagagcttctcatctacatctctcctcggaag	905
Db	31630	TCACACAGCATGGAGATAGGAGTATTATCAGAGTCTTCATCTACATCTTCCTCGGAAG	31689
Qy	906	cttgaattcatcagcaccacagtgagacgttcaactcgctgggattagtcacggtcaagt	965
Db	31690	CTTGAATTCATCAGCACCCACGTCGAGAACGTTCCAGTGAGTTTTTCCATTTTCATGCAACT	31749
Qy	966	gcgtggcaag	976
Db	31750	GTGCCCCATAG	31760

RESULT 12

AC069571

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AC069571 Homo sapiens chromosome 3 clone CTD-2015F14 map 3, WORKING DRAFT

SEQUENCE, 16 unordered pieces.

AC069571.3 GI:10140818

HTG; HTGS_PHASE1; HTGS_DRAFT.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 115766)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome 3, clone CTD-2015F14

Unpublished

2 (bases 1 to 115766)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F., Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G., Campiliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mienna,V., Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testave,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Direct Submission

Submitted (03-JUN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Sep 15, 2000 this sequence version replaced gi:9887750.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RN/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8478
Center clone name: 2015_F_14
----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 106109 bases at least Q40
Consensus quality: 111659 bases at least Q30
Consensus quality: 113401 bases at least Q20
Insert size: 121000; agarose-fp
Insert size: 114266; sum-of-contigs
Quality coverage: 3.6 in Q20 bases; agarose-fp
Quality coverage: 3.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 17190: contig of 17190 bp in length
* 17191 17290: gap of 100 bp
* 17291 18431: contig of 1141 bp in length
* 18432 18531: gap of 100 bp
* 18532 20008: contig of 1477 bp in length
* 20009 20108: gap of 100 bp
* 20109 22046: contig of 1938 bp in length
* 22047 22146: gap of 100 bp
* 22147 24142: contig of 1996 bp in length
* 24143 24242: gap of 100 bp
* 24243 27247: contig of 3005 bp in length
* 27248 27347: gap of 100 bp
* 27348 31214: contig of 3867 bp in length
* 31215 31314: gap of 100 bp
* 31315 36549: contig of 5235 bp in length
* 36550 36649: gap of 100 bp
* 36650 42234: contig of 5585 bp in length
* 42235 42334: gap of 100 bp
* 42335 45924: contig of 3590 bp in length
* 45925 46024: gap of 100 bp
* 46025 50925: contig of 4901 bp in length
* 50926 51025: gap of 100 bp
* 51026 59799: contig of 8774 bp in length
* 59800 59899: gap of 100 bp
* 59900 83502: contig of 23603 bp in length
* 83503 83602: gap of 100 bp
* 83603 92991: contig of 9389 bp in length
* 92992 93091: gap of 100 bp
* 93092 103204: contig of 10113 bp in length
* 103205 103304: gap of 100 bp
* 103305 115766: contig of 12462 bp in length.

FEATURES

source

1. .115766
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"

/clone="CTD-2015F14"
/clone_lib="CIRD Human BAC"
1. .17190
/note="assembly_fragment"

misc_feature

clone_end:SP6
vector_side:left"

misc_feature

17291. .18431
/note="assembly_fragment"

misc_feature 18532. .20008
/note="assembly_fragment"
misc_feature 20109. .22046
/note="assembly_fragment"
misc_feature 22147. .24142
/note="assembly_fragment"
misc_feature 24243. .27247
/note="assembly_fragment"
misc_feature 27348. .31214
/note="assembly_fragment"
misc_feature 31315. .36549
/note="assembly_fragment"
misc_feature 36650. .42234
/note="assembly_fragment"
misc_feature 42335. .45924
/note="assembly_fragment"
misc_feature 46025. .50925
/note="assembly_fragment"
misc_feature 51026. .59799
/note="assembly_fragment"
misc_feature 59900. .83502
/note="assembly_fragment"
misc_feature 83603. .92991
/note="assembly_fragment"
misc_feature 93092. .103204
/note="assembly_fragment"
misc_feature 103305. .115766
/note="assembly_fragment"
clone_end:T7
vector_side:right"
BASE COUNT 31539 a 24759 c 26611 g 31352 t 1505 others
ORIGIN

Query Match 3.8%; Score 266.4; DB 2; Length 115766;
Best Local Similarity 99.6%; Pred. No. 2.4e-56;
Matches 267; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 gttcgtgaaagaattgccgcttttcaaaacaaagtggaacagctcgaggaggggaac 60
|||||
Db 110015 GTTCGATGAAGAATTGCCGCTTTTCAACAAGAGTGAACAGCCTCGGAGATGGGAAC 110074
Qy 61 agagaggcgatggggctgtcagaagaatggactgtgcacagccaaggagccaccttc 120
|||||
Db 110075 AGAGAGGGCGATGGGGCTGTGAGAAGAATGGACTGTGCACAGCCAAGAGCCACCACTTC 110134
Qy 121 ggcttgagcccttccttctcctgcttggagatggagctgacacagccttctag 180
|||||
Db 110135 GGCTGGAGCCCTTCCTTCCTGCTTTGGAGATGGGAGAGCTGACCAAGCTTCTCTAG 110194
Qy 181 gaagagaaattcctcagaccagatctctcctggctgcatcttctacagacagcagcttc 240
|||||
Db 110195 GAAGAGAAATTCCTCAGGACCAGATCTCTCTGGCTGATTTCTACAGGACAGCAGCTTC 110254
Qy 241 ctctctctcttagaccttctctcacct 268
|||||
Db 110255 CTCCTCTCTTAGACCTTTCTCTCACGT 110282

RESULT 13

AF389852

LOCUS AF389852 34308 bp DNA linear PRI 23-AUG-2001
DEFINITION Homo sapiens psoriasis susceptibility gene candidate interval,
partial sequence.

ACCESSION AF389852
VERSION AF389852.1 GI:15281554

KEYWORDS
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 34308)
AUTHORS Hewett,D.R.

1 GGGAGGTGGGCTCCTGCTCATCCTAGGCATCGCACTGATTGTTACCTGTTGCAGAAGAA 60

This Page Blank (uspto)